Methods of Medical Technology Assessment with an application to liver transplantation

Methods of Medical Technology Assessment with an application to liver transplantation

Methoden voor medische technology assessment
met een toepassing bij levertransplantatie

PROEFSCHRIFT

Ter verkrijging van de graad van doctor
aan de Erasmus Univeristeit Rotterdam
op gezag van de Rector Magnificus
Prof.dr. C.J. Rijnvos
en volgens besluit van het College van Dekanen
De openbare verdediging zal plaatsvinden op
woensdag 4 september 1991 om 15.45 uur

door

GOUKE JACOBUS BONSEL

geboren te Beverwijk

PROMOTIECOMMISSIE

PROMOTORES:	Prof.dr P.J. van der Maas Prof.dr J.D.F. Habbema
OVERIGE LEDEN:	Prof.dr M.J.H. Slooff Prof.dr F.F.H. Rutten

CONTENTS

PREFACE

My scientific career as a public health researcher started in 1985 as co-investigator in the two medical technology assessment studies described in this thesis. The concept of this thesis was born during my application interview. Upon a question of the director of the institute of Public Health during the application interview, whether I intended to write a thesis on the studies to be conducted, I showed him some reluctance to do so: better to judge all the evidence afterwards.
The judgment turned out favourably, but the writing took a bit longer. Nevertheless I am happy with the course of events. Since the final reports on heart and liver transplantation were published I increasingly became aware of the common themes in this type of research.
Paul van der Maas provided me the opportunity to write this thesis as a scientific reflection on medical technology assessment, rather than a summary of empirical articles, which of course are relevant for their own sake.
The reader should be warned: the text is rather demanding. Chapter 3 is the nucleus, proposing a methodological framework of medical technology assessment. Any reader who aims at reading at least a part of the thesis, should not skip this chapter. Chapter 1 is a prelude and chapter 2 a reflection on medical technology assessment and health policy. The last three chapters on survival, quality of life and need for medical services are rather technical. They elaborate on three issues which have constituted the major part of my empirical work.

VOORWOORD

Mijn wetenschappelijke loopbaan startte in 1985 als mede-onderzoeker in de twee evaluatie-onderzoeken, die in dit proefschrift zijn beschreven. Het achterliggende idee van dit proefschrift werd tijdens mijn sollicitatiegesprek geboren. Op een vraag van de directeur van het instituut, of ik voornemens was de uitkomsten van het uit te voeren onderzoek tot een proefschrift te bewerken, reageerde ik afhoudend: verkoop de huid niet voor de beer geschoten is.

Met die beer kwam het wel goed, maar het schrijven van het proefschrift nam meer tijd in beslag. Niettemin ben ik blij met deze loop der gebeurtenissen. Sinds het verschijnen van de eindrapporten over hart- en levertransplantatie ben ik mij in toenemende mate bewust geworden van de rode draad in dergelijk onderzoek.

Paul van der Maas bood me de gelegenheid dit proefschrift in deze vorm te schrijven, namelijk als een wetenschappelijke bezinning op medical technology assessment, een term die mijns inziens niet bevredigend kan worden vertaald.

De lezer zij gewaarschuwd: de tekst is weerbarstig, waarbij de Nederlandse lezer nog gehandicapt wordt door het feit dat de tekst naar nieuw academisch gebruik niet in het Latijn maar in het Engels is gesteld. Hoofdstuk 3 vormt de kern; hierin wordt de methodologische structuur van medical technology assessment uit de doeken gedaan. Hoofdstuk 1 is een prelude en hoofdstuk 2 bespreekt de verhouding van technology assessment tot het gezondheidsbeleid. De laatste drie hoofdstukken over overleving, kwaliteit van leven en de behoefte aan medische technologie zijn nogal technisch van aard. Zij werken de drie onderwerpen uit waar ik mij in empirische zin het meeste mee heb beziggehouden.

1 INTRODUCTION

1.1 The general and the specific

Medical care has changed dramatically in recent decades in terms of volume, depth and pace of change. This development may be observed from many viewpoints and may be analysed by a variety of methods.

This thesis concerns a recently developed approach of analysing the changes in medical care, known as 'medical technology assessment'. Essentially its application should yield information to support decision-makers at any level in the health care system. The results of our experience in the development and application of medical technology assessment are presented in two parts.

The first part gives a general theoretical overview of the subject. It includes a description of the particular context in which medical technology assessment (MTA) is used with an account of the methodology with emphasis on three topics of medical interest. The last chapter of this part discusses some conclusions and general recommendations.

The second part is added as an appendix. It embodies the empirical part of the thesis and demonstrates the results from practical applications of the proposed methodology. Most evidence is drawn from the MTA of liver transplantation, which started in 1985 and presented its final report in 1988. Many examples in the first part of this thesis are also taken from this study.

Although liver transplantation is a medical technology with modest impact on health care and on overall population health, and although the MTA was primarily commissioned by health insurance parties to support insurance coverage decisions, we assume this case to be representative for MTA as a general approach to changes in health care.

This introductory chapter starts with a general description of the pro-

blems related to changes in health care (section 1.2) and the important contribution of new technologies to these changes (section 1.3). The search for information on new technologies will be explained as a natural reaction of all parties involved towards a changing environment (section 1.4). It also reflects the growing awareness of the economic relevance of this information.
Special attention is paid to the position of health care authorities (section 1.5). There has been a noticeable change in their attitude. Decisions concerning the introduction and diffusion of new medical technologies are no longer domain of professional interest alone (Freidson, 1972). Even the routine application of existing technologies has become a field of societal interest, particularly if variations in practice suggest the waste of precious resources (Wennberg, 1973; McPherson, 1990).
In addition an increased awareness of the scarcity of resources may well be observed (Relman, 1988). Health care authorities, hospital managers and other decion makers are developing strategies for controlling the diffusion of emerging technologies. This control is usually based on supportive information about various aspects of the technology.
The response of Dutch health care authorities to the introduction of liver transplantation is summarized (section 1.6). Medical technology assessment was part of their strategical response.
In section 1.7 we formulate three key questions which we regard to be relevant in the judgment of the current and future value of MTA.

1.2 Changes in medical care

On the level of *medical science* per se, an impressive increase of fundamental and applied knowledge can be observed. Scientific progress in

immunology and molecular (patho)biology is just one example with high public appeal.

Medical practice has changed too, not only with the introduction of new drugs, devices and equipment but also with concomitant changes in the delivery of medical care. The expansion of knowledge is mirrored by the fragmentation of the medical profession into highly specialized subdisciplines, with consequences for every level of care. The organizational complexity of a transplantation programme is just one illustration of these consequences. The hospital has gradually changed from an institution for supportive care towards an institution for diagnosis and, where necessary, acute treatment[1]. As a consumer, both from an economic and a philosophical point of view, the patient has entered medical practice. The objectivization of the craft, though essential to success in complex treatments, has given rise to alienating effects. This alienation has clearly provided opportunities for a growing number of alternative healers to cultivate the nostalgia for 19th century paternalistic and ineffective medical practice with a flavour of naturalism.

Medical society, too, has been affected by new developments. Inevitably the art to care, cure, detect and prevent has been redefined, reversing this order of interest. Specialists are facing shifts in the borders of their territories, ambivalent about the choice between further fragmentation and only partial mastery of their craft[2]. The introduction of expensive new technologies has given rise to visible concern about the consequences of limited availability of treatments.

The *medical system* has also changed, providing - in Western societies[3] - the opportunity for an increasing number of consumers to profit from increased medical supply. This change can be observed regardless of the system of health care insurance, though different distributional effects can be observed from particular systems.

The relation between *society and medicine* has been affected by changes in both domains. For example societal developments, like individualization and women emancipation, have entered medical knowledge and professional practice. On the other hand, the ever-increasing potential of medical science has given rise to high expectations of the power of medical professionals to solve problems of many kinds.

We should be aware that this description has a critical counterpart.

Medical science has grown exponentially, but the increased domain of 'the known' seems to result in a greater domain of 'the unknown'. The scientific success of molecular medicine is beyond criticism and has strengthened its paradigmatic position. However, questions remain with regard to its blessings for the population at large. Major disease categories like cardiovascular disease, many malignancies and an array of infectious diseases continue to scourge society (Van der Maas, 1989).

The increased Taylorism within *medical practice* obviously produces efficient care at an optimal level, but its disorienting effects cannot be ignored. One of the attractive aspects of alternative healers is their independence from fragmented diagnostic and therapeutic modalities and their simple definition of pathophysiology of disease, blaming the patient if disease occurs or progresses. The eye of God may be preferred to blind biological chance.

Professional developments should also be regarded from a critical point of view. Medical training continues to be a rather closed shop. Socialization still dominates practical training (see e.g. Fahrenfort, 1985 [especially p.179-]) and implicit variables are decisive for the acquisition of a specialist training position. Even more disturbing is the persistence of a modern form of slavery, forcing many in search of a specialist training position to work under sometimes extremely unfavourable conditions.

Some footnotes may be added to the blessings of the *medical system*. The claim of social justice through universal acces to medical care can not be upheld. Obviously granting equal rights does not imply automatically a fair distribution of medical care (Wagstaff, 1989).

Finally, from a critical point of view, the intrusion of medicine into everyday life as a *societal* hazard should be mentioned. This intrusion is only partially covered by the connotations of the word 'medicalization'. Maybe Foucault's philosophical vehemence has best described the invading potential of medical power (Foucault, 1976 [p.76-], 1980).

1.3 Technologically induced changes

Emerging and changing technologies play a dominant role in many changes (Reiser, 1978; Showstack, 1982; Hoffenberg, 1988)[4]. Before 1900 doctors were long on charity and short on science (Jennett, 1986 [p.22]; McKeown, 1973), but this picture has changed. Below we can only give an impression of the role of technologies, applying the same classification as in the previous section.

Medical science clearly benefits from technological advances in basic sciences as well as in applied and industrial sciences. Manufacturers of drugs and medical equipment clearly add to the opportunity to disseminate innovations by their ability to upscale production of succesful technologies quickly (e.g. the rapid appearance and spread of several cyclosporin determination kits).

Medical practice has changed the last decades with the introduction of many succesful drugs, e.g. analgesics, betablockers, anesthetics, antipsychotics, diuretics, steroids and other hormonlike drugs, with H2-blockers and cyclosporin as two recent and succesful examples.

Many devices have changed the hospital environment, e.g. the SMAC, monitor equipment, a multitude of diagnostic imaging devices and supportive systems which may temporarily replace vital organ functions. Many of these devices (e.g. the veno-venous bypass; Slooff, 1989) are also involved in liver transplantation.

Finally the *organization of medical care* has been affected by the management and the information sciences (Greer, 1977; Jennett, 1986 [p.19-, 71-]). Traditional organizational settings have been replaced, and new organizational solutions are offered to otherwise unresolvable problems, such as in population screening, but also in organ transplantation.

Some specific problems within the *professional societies* caused by new technologies should be mentioned. New diagnostic techniques raise questions about the specialism most suitable to perform the test (see e.g. ultra-sound imaging techniques). Within the medical society questions arise when exogeneous factors limit the number of providers of a specific treatment modality, as was the case in liver transplantation.

Related problems formerly sophisticated technologies like ECG, immunological tests, pregnancy tests, infusion devices become available to primary care physicians and in some cases laymen. The care of insuline-dependent diabetics, for example, has been radically changed with the availability of recombinant insuline in 'leg'-friendly administration pens and the use of autotest devices for self-regulation.

Technological change, including this development of user-friendly devices, continues to blur traditional inter- and intraprofessional borders and creates new areas of professional interest. Technological advance is also partially responsible for a shift in care-delivery towards less structured environments. Home care probably will probably show an exponential growth, as many supportive systems (drugs, information, transport) have become easy to operate and reduced to manageable dimensions.

The *health care system* must find an answer to emerging technologies that do not fit easily in its framework. Given national variations in medical systems, national health authorities and health insurance companies do not always face the same problems (McPherson, 1990). However, a general pattern in the 'getting noticed' process may be observed.

Table 1.1 Factors influencing the explicit reaction of the health care system to an emerging medical technology

1. the degree to which national regulations, reimbursement schemes and professional conventions determine medical practice with regard to the technology concerned (where, for whom, by whom);
2. the degree of diffusion of the technology;
3. the degree to which the technology requires new, expensive equipment, or existing equipment to be transferred from routine work; the same for human resources;
4. the degree to which the technology requires a new organizational structure or substantial changes of existing ones;
5. the degree to which the technology is thought to have high societal impact (e.g. technologies applied at the beginning and the end of life).
6. the degree to which registration of a new technology is necessary for introduction on the health care market.

An emerging technology will force the current medical system to formulate an answer if on one or more of the following reaction-provoking circumstances is relevant (see Table 1.1). Resource scarcity (underlying 1., 3. and 4. of Table 1.1) is not the single important factor. In most Western countries registration obligations and ethical considerations (5.

and 6. of Table 1.1) are important as well[5]. Most, if not all of these factors also pertain to the liver transplantation case in the Netherlands.

Finally the *society-medicine* relation is influenced by the diffusion of existing and the development of new technologies[6].

Technological advances have redefined social relations. For example the relation between reproduction and sexual activity has been dissolved, exerting influence on partnership at the legal, social and individual level. Biological and social parentship are no longer tautological. Supportive systems for the period of conception and premature birth are increasingly reducing the period of obligatory fetal residence in utero.

The growing potential of genetic information on the proneness to disease will not only affect decisions about reproduction but more generally affect decisions about employment, insurance and mortgage arrangements. At the other end of the life span, supportive systems (technical and organizational) provoke questions about a suitable living environment for the elderly. Advanced age is no longer an absolute contra-indication for complex surgical procedures (e.g. coronary artery bypass grafting, liver transplantation) or life sustaining systems (e.g. hemodialysis). Generally technological advances have raised questions about their use in terminal patients.

1.4 Information on changes in medical care

As technological advances usually induce changes on many levels of the health care system, they are accompanied by an array of reactions.

The intensity of this response is primarily determined by the recognition of change[7] and the perception of interests, violated or supported by the application of the technology concerned[8].

In Table 1.1 we already listed some factors contributing to *recognition* of change. The interests involved may range from the financial to the ethical, though the former seems the more stimulating. We think that simple positive technological changes, which serve the interests of all, are uncommon. A mix of positive and negative effects is more likely. Hence the reactions will depend on the specific position of the actor in the health care system.

Irrespective of the position involved, *information* on the new medical technology plays a pivotal role in the reactions of interested parties, and consequently in the process of technology diffusion (Greer, 1977; Casparie, 1989 [p.17-]). As a rule, this information is used as an instrument for decision-making, either directly or indirectly, depending on the party involved (IOM, 1985 [p.26, 32-69]). A sample from parties with interests in liver transplantation may substantiate this view.

The liver transplantation team needs information to support patient management, to acquire the cooperation of in-hospital workers and of the hospital management, and to provide out-hospital colleagues with information on process and outcome. The hospital management needs information on the resources devoted directly or indirectly to the liver transplantation programme, and on the organizational impact and the friction problems accompanying its introduction[9]. The patients need information to support their treatment decision and to guide their choice for the best environment to receive this treatment. They should e.g. know whether the technology involved has been studied in patients similar to themselves. The health insurance company is interested in the potential benefits, costs and savings, but also in information on structure and process of the programme. Costs and benefits may support the decision whether insurance coverage should be extended to liver transplantation; process and structure information provides cues to an effi-

cient deployment of available resources. This description of the informational need with a short reference to its purpose can easily be extended to other parties (see Table 1.2).

Table 1.2 Organizations with informational need on emerging medical technologies.

Government	Health Care Providers
Department of Health	Professional Organizations
Health Insurance Companies	Drug Manufacturers
Scientific Advisory Boards	Equipment Manufacturers
(Other) Health Care Regulating Authorities	Hospital Administrators
	Market Research Companies
	Investment Companies
Health Care Consumers	
Consumer Organizations	Medical Schools
Patient Organizations	Medical Journals
News media	Public Health Scientists

This need may be satisfied in several ways, ranging from a series of phone calls to medical opinion leaders or an excerpt of published literature to a comprehensive empirical study as is described in this thesis. We shall focus on the response of health care authorities towards new technologies, and on the role of information within this response.

1.5 The structuring of the public debate on impending medical technologies

Technologically induced health care changes have given rise to growing concern about societal side effects[10], some of them mentioned in section 1.2. The uncontrolled character[11] of these changes have changed the attitude of scientists, health care providers, consumers and health policy makers (Jönsson, 1990). The unfruitful moral debate about the a priori 'good' or 'bad' character of technology has gradually been replaced by a structured, scientifically and politically more mature discussion between the various parties involved [12,13,14]. For an account of this change of battle see Cochrane and Jennett (Cochrane, 1973; Jennett, 1986 [p.169-174]).
Consensus has grown on the desirability to subject some of the new technologies to structured, explicit and public decision making[15], for example by imposing a specific registration requirement. Health care authorities have accepted their role in structuring this public debate.
Although different interests may underlie the willingness to cooperate in this process (IOM, 1985 [p.26-30]), this convergence of attitudes can be discerned in most Western countries, including the Netherlands (Geleijns, 1990; Rigter, 1989 [p.39-43]). The process in the different countries has some national pecularities owing to the structure and the mores of the health-care and the political system[16]. However, a general pattern may be observed, and will be described in chapter 2.

An important feature of this process is the explicit role assigned to information, mainly quantitative information, on a number of issues. As scientific literature can only partially satisfy this information need, additional information must be collected. Specific methods for information generation usually have to be applied. This systematic information generation to support societal decisions on medical technologies has conveni-

ently been called medical technology assessment (see Banta, 1990 for a historical review).

This thesis is dedicated to one modality of MTA, which will be called *empirical medical technology assessment*.

Next we give a summary of the response of Dutch health care authorities to the introduction of liver transplantation. Empirical medical technology assessment was a critical part of this response. As this case of MTA will serve as an example throughout the thesis a brief description seems justified at this place.

1.6 The empirical MTA of liver transplantation in the Netherlands

In the Netherlands empirical MTA as a support to decision-making was introduced in 1985 by the National Health Insurance Board (NHIB). This initiative resulted from the growing concern about the uncontrolled extensions of the coverage of the national health insurance scheme[17]. Three technologies were chosen as to be subjected to MTA: liver transplantation, heart transplantation and in-vitro-fertilization. This NHIB initiative was supported by other Dutch health care authorities and private health insurance companies. Two research groups were put in charge of the required empirical MTA's (Erasmus University Rotterdam: heart transplantations, liver transplantation; State University of Limburg: in-vitro-fertilization). As the author of this thesis was particularly involved in the liver transplantation MTA, we shall proceed with an outline of over this study.

In the Netherlands orthotopic liver transplantations have been performed since 1979 at the Academisch Ziekenhuis Groningen (AZG)[18]. The transplantation programme at the AZG started as an experimental

clinical programme, funded by general research grants. Despite recommendations of the Consensus Development Conference of the National Institutes of Health in 1983 and the favourable development of the programme, Dutch health care authorities were reluctant to incorporate liver transplantation in regular reimbursement schemes.
As a result the AZG received supporting grants from the Dutch health insurance companies to continue their programme on a larger scale from 1985 onwards, one of the conditions being that they should participate in a medical technology assessment (MTA) by an independent research team. This MTA was to yield objective information about costs, effects, organizational aspects, donor supply and other topics, which health care authorities regarded necessary for a well-founded decision. The final report appeared in September 1988, describing the experience of 221 screened and 73 consecutively transplanted patients (Habbema, 1988). The results were used in the subsequent decision process (*A1, A10*). Based on the information of the final report it was decided that the restrictive regime, limiting liver transplantation to one national centre, should be continued while additional research should be done on information about long-term results of liver transplantation (NHIB, 1988).

1.7 The aim and the objectives of the thesis

From the experiences of the Dutch empirical MTA's of heart and liver transplantation, it was concluded that empirical MTA may be very rewarding as an information generation tool (*A10*; Van Rossum, 1990; NHIB, 1990). In further elaboration of this conclusion the aim of this thesis is to provide answers on the following questions about empirical MTA in *general* (see next page):

1. What is the scientific basis of an empirical MTA?
   More specifically, what are the features of survival analysis, analysis of health status (or quality of life) and assessment of need?
2. Which conditions determine the fruitful application of empirical MTA results?
3. How can MTA be developed further?

The first question is dealt with in chapters 3 to 6. In chapter 3 a standard format of empirical MTA is proposed. In chapters 4 to 6 we elaborate on methodology for survival analysis, quality of life analysis and analysis of need, three of the core issues of an empirical MTA.

Answers to the second and the third questions are to be found in chapter 2, 3 and the appendix. In chapter 7 the answers to the three questions are summarized.

Illustrations of the proposed methodology can be found in the appendix. Appendix 1 presents the overall approach of the Duth MTA of liver transplantation, focussing on the relation between the results and their subsequent use. In appendix 2 the results of the application of cost-effectiveness analysis to the same data is shown. The following appendices are devoted to specific methodological issues of MTA. Appendix 3 shows the results of a new mathematical technique for estimating survival if transplantation is not applied. Appendices 4 to 8 all deal with different aspects of quality-of-life analysis in MTA. Appendix 4 describes the general approach. Quality of life description in liver transplantation is demonstrated in appendix 5. Three related investigations on normative quality-of-life valuation are shown in appendices 6 to 8. Appendix 9 shows the results of three techniques for estimating the need for liver transplantation. Finally appendix 10 evaluates the burdens and benefits of the three Dutch MTA's which started as a test-case in 1985.

2 THE POLITICAL CONTEXT OF MTA

2.1 Introduction

This chapter presents a prototypical description of the political context of MTA. This description at least partially fits current health policies which aim at information-supported societal decision-making on medical technologies (Jönsson, 1990; IOM, 1985)[1]. The policy process may be regarded as a decision *cycle*, with alternating phases of definition, measurement, valuation and control (see Table 2.1)[2]. The roles of health care authorities, MTA-investigators, and clinical experts vary accordingly.

Table 2.1 The decision cycle on new medical technologies.

Phase 1	Recognition and selection of technology
Phase 2	Definition of the decision problem
Phase 3	Definition of the information need and selection of the information generating modality
Phase 4	Information generation and presentation
Phase 5	Valuation, decision-making and implementation
Phase 6	Monitoring and repeated assessment

This chapter elaborates on these phases and will be ordered accordingly. In the notes the equivalents of the proposed phases in the current Dutch health care system will be presented.

2.2 Phase 1 - Recognition and selection of technology

The initial stage of the decision cycle addresses the following questions:

a. how do new and existing technologies which deserve reconsideration get noticed (recognition)?
b. how may technologies which lend themselves for MTA be selected (selection)?
c. who is put in charge of the recognition and selection (execution)?

Recognition depends on many factors (see Table 1.1). From this table follows that passive recognition can be supplemented by an active search for technologies to be subjected to MTA.

The process of recognition will also depend on a technology's stage of development. New technologies which deserve attention may be detected by screening medical literature, by visiting medical congresses and by regular consultation of experts. Some have argued in favour of 'early warning systems' to obtain timely knowledge of technologies which are on their way in (Banta, 1987). The arguments rest on the assumption that so far shortage of time has prevented conscious decisions being made and concommittant actions undertaken on further application of a technology. We do not think that this assumption is justified. Besides, these systems are likely to take up considerable resources[3,4].

Existing technologies which deserve reconsideration may be found in a similar way. Analysis of geographical variations in medical services is an additional possibility (McPherson, 1990; Phelps, 1990; in the Netherlands Mackenbach, 1990 and Anonymous, 1990)

'Getting noticed' does not imply that a technology is automatically a candidate for MTA. Additional selection is necessary based on a priori assessment of the relevance of societal decision making and on judgment of the additional value of high quality information[5]. Criteria to judge relevance may be developed[6]. In the Netherlands it would seem appropriate to base these criteria on the five boundaries of health care as

defined by Dutch health care authorities[7]. Notwithstanding national pecularities, there appears to be an international consensus on which medical technologies should be subjected to MTA[8].
The last issue is the assignment of roles in the recognition and selection process. The diversity of interests involved and the relatively unstructured appearance of new medical technologies (see chapter 1), in our view precludes a prescriptive answer to the question of how recognition should take place. For maximum sensitivity and involvement, access to the recognition process should not be restricted to those parties directly involved in the later formal decision process[9,10]. Selection already includes the weighing of societal interests and its realization should be treated accordingly[11].

Phase 1 concludes with the overall judgement of whether continuation of the cycle is justified[12].

2.3 Phase 2 - Definition of the decision problem

In Phase 2 the decision problem is framed[13]. Usually the perception of the decision problem is rather straightforward: should society give a new technology legal, financial and organizational support? We think this perception is a simplification which is usually not appropriate.
A careful definition of the decision problem is thus required, consisting at least of the following three parts.

First, the definition of the *subject* of the decision[14]. The technology should be described and the scope of decision explicitly stated. The technology description includes a description of its physical nature, purpose and organizational context (see Table 2.2 on next page).

Table 2.2 Descriptive attributes of a technology

Physical nature	
Drugs and chemical substances for other use	Medical procedures
Devices	Surgical procedures
Communicative procedures	Medical support systems
Information systems	Organizational systems
	Other
Organizational setting	
Home	Hospital
General practice	Specific units
Outward clinic	Other
Primary objective	
Prevention	Rehabilitation
Diagnosis	Other
Treatment	

The scope of the decision concerns questions such as 'To what degree should the decision be extended to existing alternatives?', 'Should the decision give a simple, unconditional 'yes or no' answer according to the above-mentioned perception, or should conditions be included (for whom, by whom[15])?'. The last question will generally be relevant.
The scope of the decision will differ with the health care party involved. A particular hospital manager will be interested in whether he should

start supportive or restrictive actions within the hospital environment. A health insurance company will consider its portfolio of insurance arrangements and reimbursement conditions. The scope of national health care authorities is broad, as it not only deals with public health considerations, but also with the interests of other parties involved.

As a second part, the issues thought to be relevant should be defined. To this description issue-specific *criteria* to be used in Phase 5 may be added[16]. A priori (tentative) definition of criteria should be aimed at. The translation of issues, which are usually theoretical concepts like 'effectivity', 'safety', 'real need', to the language of empirical observations is not that simple, as will be shown in the following chapters.
As with the scope of the decision-problem, the perspective of the decision-maker will influence the issues to be taken into account.
Finally, the *context* of time and place should be described; is the scope local, national or even international, what is the timetable of decisions and, eventually, implementation and re-evaluation.

The decriptions in Phase 2 will be used as a base for information generation decisions in the subsequent phases.

2.4 Phase 3 - Definition of the information need and selection of the information generation modality

In Phase 3 the information need must be defined and the preferable method of information generation determined.

The information need is theoretically defined by the aspects and the associated criteria distinguished in Phase 2. Intangibles form a specific category of issues, particularly if they are considered to play a major role[17]. The information need may not easily be defined in this case.

In some cases definition will be straightforward and information sources will be readily available, resulting in a quick generation of information. However, in a typical situation, definitions in Phase 2 and 3 are rather vague and information will be patchy.

Existing sources of information will first be screened for the required information. With new technologies, available literature will usually only partially satisfy the information need. The institution of a specific information generating process like MTA may then be considered.

Three prototypes of MTA can be distinguished, meta-analysis of literature, Delphi-techniques and empirical analysis of raw data respectively. In the latter case data has to be extracted from a clinical pilot programme. Crude data on existing clinical programmes, even if they exist, will not be easily made available[18]. The essential features of these methods will be summarized in section 2.5.

Available literature should be scrutinized first as any additional data-collection implies considerable investments (Olsen, 1982; *A10*). Representativeness, comprehensiveness and comparability determine the suitability of existing literature (NHIB, 1986 [p.56], Bonsel, 1988A).

With regard to representativeness the disparities between the Netherlands and the United States, the major source of new technologies and corresponding literature, will frequently limit the suitability of existing information. Clinical practice frequently differs. The role of surgery vs. internal medicine (hepatology, cardiology, nephrology) in solid transplantation is but one example. More generally the medical system differs for example with regard to referral practice and health insurance arrangements. In this context the different economies of health care are quite important. See e.g. the remarkable variation in medical purchase power parities (Iglehart, 1988). These variations are also likely in solid organ transplantation (Krüger, 1989; Health Council, 1989 [p.124]).

Comprehensiveness pertains to the presence of information on all core issues. With new technologies attention is usually confined to safety and efficacy. Information on other topics is available only infrequently. Hence additional research usually is necessary if a complete MTA as defined in chapter 3 is aimed for.

Comparibility pertains to the use of standardized methods to collect and present information. Three examples are given of implicit differences in definition of methods, definition of patients and choice of measurement instruments, which affected comparibility.

In liver transplantation the calculated survival of transplant patients sometimes apparently excluded 24-hour peri-operative mortality. As peri-operative mortality may account for about 30% of total (5-year) mortality, comparison of various centres was not feasible.

In liver transplantation the use of published survival information was also limited due to different patient selection criteria. At the start of the Dutch MTA in 1985, the literature on survival effects of liver transplantation was limited to reports of survival following liver transplantation. Quantitative information on the severity of liver disease at the time of transplantation was lacking but it was generally accepted that the centres were heterogeneous with regard to pre-transplantation status. Hence the net effect of transplantation could not be extracted from the literature and different centres could not validly be compared (Bonsel, 1988F).

The last example is drawn from the heart transplantation study. The use of a wide array of unique quality-of-life instruments in previous studies disabled the comparison of results on this issue (Bonsel, 1988A,C). The issue of standardization will be further elaborated on in chapter 3.

If the information need is not fully satisfied by existing literature, an empirical MTA should be considered. Justification should take marginal information, investigational costs and time constraints into account. The

concepts of effectivity space and cost space may be helpful (Drummond, 1987 [p33,34]; *A10*)[19]. Obviously an empirical does not exclude the use of literature and experts to provide a.o. reference values and valuations.

Phase 3 is the starting point of the interplay between health policy makers, health care providers and scientists who might be put in charge of information generation. If a separate investigative effort is aimed for (one of the three MTA-modalities), this phase contains the design of that study and of the associated clinical protocol[20,21].

2.5 Phase 4 - Information generation (MTA strictu sensu)

The three modalities of information generation will be summarized next.

Meta-analysis is the formal synthesis of available information or data on a particular topic by means of standard statistical procedures (Van Houwelingen, 1989 [p.184-192], IOM, 1985 [p.125-128]). It has its origin in the social sciences (Glass, 1981), but its value in counting the evidence of numerous 'small-difference'-trials in medicine is currently acknowledged. Its primary use is in outcome-analysis, enhancing the discriminative power of a number of comparable trials, while providing information on the generalizability of their conclusions.

The Delphi-method is based on a formal interactive survey process (IOM, 1985 [p.130-131], Casparie, 1989 [p.180, 181]). The method generates information from the responses of mutually anonymous experts, applying controlled feedback procedures. The procedure starts with obtaining individual opinions by some formal method (questionnaire, structured interview). Next, data-enhancing techniques are applied (semi-quantification and aggregation). Each expert is confronted with his previous response and with the distribution and/or aggregation of the

group response. In a cyclic process, convergence or divergence is aimed for, maximizing the yield of the exact and intuitive knowledge of all experts. The method is used to obtain predictions of events, trends and of (quantitative) estimates when empirical data are scarce. The Delphi-method may be a particularly suitable device if technologies are still too immature to justify widespread use in humans.

Empirical MTA is the formal analysis of the effects, costs, and future impact of a technology based on data abstraction form current clinical research programmes. A standardized methodology is used (chapter 3).

Meta-analysis and Delphi-studies will not be discussed further. For an example of multiple meta-analyses, see Keirse (1988). For examples of Delphi-studies, see Van Beeck (1989) and Milholland (1973).

Policy makers play only a modest role in this phase. In the case of an empirical medical technology assessment this role is primarily the optimization of conditions to carry out the proposed investigation[22].

2.6 Phase 5 - Valuation, decision-making and implementation

Phase 5 consists of weighing various aspects, the decision and its subsequent implementation[23]. The political undesirability[24] of MTA results (e.g. due to expected difficulty with the implementation of decisions) should not influence decisions to be taken too much.

Decision-making in the Netherlands regarding liver transplantation has been described (Van Rossum, 1990; *A1*, *A10*). De Charro (1986) provided an account of Dutch health policies on transplantations in general. From these papers it follows that MTA results usually are relevant for all parties involved, supporting various decision processes.

2.7 Phase 6 - Monitoring and repeated evaluation

In this phase the technology is monitored; this should take place[25] in the case of one of at least three situations.

A decision to *halt* the diffusion of a technology justifies monitoring of further scientific advances. The information may be retrieved from the literature and from carefully selected experts at regular intervals.

A *conditional positive* decision may need reconsideration after some time. In this case monitoring the application of the new technology may provide the required information. Reconsideration might result in the relaxation of the constraints (by whom, for whom), but also in modification of the budget in view of scale and learning effects.

Even the monitoring of a medical technology with an *unconditional positive* decision on its application may be justified if uncritical extension of its application may be expected (Anderson, 1988)[26].

2.8 Summary

In this chapter a particular response to technologically induced changes was described. A cyclic structure was presented with alternating phases of definition, measurement, valuation and control. Although the point of departure was the position of health care authorities, this description, with some adaptations, might also fit the decision-processes of other actors. We are aware that other responses are conceivable.

In the following chapters the methodology of empirical MTA's will be discussed from the viewpoint of societal decision-making. Hence we focus on one specific modality of Phase 4 of the proposed decision-cycle.

3 EMPIRICAL MTA: ISSUES AND CONCEPTS

3.1 Introduction

In this chapter we shall present an outline of the methodology of empirical MTA (empirical MTA is further usually indicated by MTA sec).

Section 3.2 presents a standard format of MTA.
The same section discusses briefly other aspects of methodological standardization, particularly the use of randomized controlled clinical trials (RCCT's) and the application of quantitative models. The role of the perspective of the MTA - so far we have adopted the perspective of society at large - is also indicated.

Five issues are considered essential in MTA: the functioning of the health care programme in which the technology is embedded, the effects on patient outcome, the costs of the use of the technology or its alternatives, the need for the technology, and current and future uncertainties. Sections 3.3 to 3.7 introduce these issues, with an emphasis on conceptual considerations. Each section ends with an indication of how particular types of technology affect the empirical task.
Two of these issues (the effects on patient outcome - survival and quality of life - and the need for the technology) are described in more detail in chapters 4 to 6.

The last section of chapter 3 is devoted to two technical features of MTA and two additional topics which may be included in an MTA i.e. distributional analysis and legal/ethical analysis. Social issues (see e.g. IOM, 1985 [p.154-158]) are included in the issues mentioned above (see note)[1].

3.2 Core issues in MTA; other aspects of standardization

The standard format of an MTA

From a societal perspective, the following set of issues seem relevant (see Table 3.1, derived from IOM, 1985 [p.34, 153], Habbema, 1989).

Table 3.1 Standard format of an MTA from a societal perspective

Core issues
Action of the technology
Effects on patient outcome
Effects on the use of resources
Need for the technology
Current and future uncertainties
Optional issues
Distributional effects
Legal and ethical aspects

Published MTA's suggest that there is indeed international consensus about the issues which are critical to societal decisions on a medical technology (see e.g. White, 1982 [p.158]). Distributional analysis and legal and ethical considerations are regarded as optional issues. Therefore most attention will be given to the five core issues.

The need for a standard approach to the issues

Available methods may provide us with an empirical approach to most of the issues in Table 3.1. However, standardization of these approaches is needed for four reasons:

a. to enable comparisons of the same technology applied in different circumstances (patients, settings, countries);
b. to enable comparisons of different technologies for different purposes;
c. to enable full economic analysis as part of an MTA;
d. to enable integral forecasts of the health-care programme.

Comparibility requires universal concepts, and reproducible and generally applicable instruments to measure them, with a protocol for their standard application. Then the results of an MTA can be compared with the results of MTA's of the same or quite different technologies.
The presence of a standardly derived global set of reference values of health care sectors (Koopmanschap, 1991; Health Council, 1986 [p.122-123]), makes it possible to select the best health-care programme for a given health problem. Moreover, comparability with overall performance of other programmes will assist the decision-maker on the global level.
Full economic analysis (or economic evaluation) is described in section 3.5. Theoretically it is a powerful instrument for resource allocation in health care. Even if resource scarcity is not the main reason for the institution of an MTA, economic evaluation is regarded as indispensible in MTA to enable an optimal choice. It requires a quantitative approach to the concepts and operationalizations of input, throughput and output of a health-care programme (Drummond, 1987A; De Charro, 1989).

A specific feature is the requirement of rigorous analysis of the target programme and of its best alternatives. Finally it usually includes the computation of aggregate measures of economic desirability (e.g. the cost-effectiveness ratio).

The requirement of rigorous analysis of the best alternative is conceptually closely linked to the experimental method (in this context usually an RCCT) which is applied to get unbiased estimates of relative effectiveness of a new technology[2,3]. Although this experimental design is not always possible, it is undulty ignored on many other occasions[4].

Forecasting essentially requires a quantitative model, which approximates the 'real world' with regard to its essential features. In our case the model should consist of a dynamic disease model which is linked to a technology application model using the same concepts of input, output (consequences on among other things the disease level) and time of occurrence. A concise report on the key position of quantitative modelling in health policy is found in Van der Maas (1986). Eddy lists some technical essentials (IOM, 1985 [p.153], see also Eddy, 1980). Elaborate examples are included in Van Oortmarssen (1990) and Van Hout (1990).

Thus we arrive at the following standard elements of MTA:

a. a dynamic disease model linking the occurrence and the development of disease to survival, health status and other events amenable to intervention;
b. a technology application model, linking the employment of resources to the disease model via defined interventions;
c. an overall structure of the model which allows for comparison over time of alternatives;
d. uniform concepts for input, throughput, output and need, which are interchangeable in all the branches of the model;
e. uniform non-disease-specific measurement instruments for output

measurement; uniform non-setting-specific measurement instruments for input measurement;

f. a comprehensive approach to uncertainties.

The perspective of the MTA

Other perspectives than that of society may be conceived of (see Table 1.2, section 2.3 and Drummond, 1987 [p.7, 19]). Luce distinguishes three additional viewpoints: third party payers, health care providers and patients (1990 [p.45]). Though some heterogeneity may be found within these classes[5], this distinction may be useful (see e.g. Weisbrod, 1980).

A societal perspective is generally regarded as the most comprehensive (Drummond, 1987 [p.39-40]). The structure of an MTA from other perspectives may usually be derived from the structure shown above (Luce, 1990 [p.46-49]; Drummond, 1987 [p.39-40]). However, other perspectives may imply that the contents of issues will be adapted e.g. by changing the emphasis on particular topics, by applying other valuation principles or other aggregational techniques. Here the influence of other perspectives will not be systematically addressed.

3.3 Action of the technology

Relevance

The relevance of this issue is twofold. First, it provides information on the subject of the decision, which also defines the boundaries of the MTA. It provides also insight in the process of technology application.

A comprehensive description of the action of the new technology is usually not available from the clinical protocol. For example, often the protocol does not include a detailed description of the selection process and of the available alternative options. As a result the decision-maker faces a decision problem with apparently one option. Apart from support to the decision-maker in that case, a detailed description is supportive for the MTA-analysis of other issues. A detailed description of selection criteria of patients is e.g. required in the analysis of outcome and of need.

Information from the process of technology application may be useful for evaluation of the functioning of the protocol. It may also provide clues to the improvement of e.g. organizational and legal arrangements.

Methods

The qualitative part of the description includes:

a. a description of the technology and its mode of action;
b. a description of the purposes of its application;
c. a description of the patients to be subjected to the technology;
d. a description of the environment of the technology;

Below we give two additional remarks on b. and c.

The purposes of application may be defined on the process and the outcome level. From an MTA viewpoint purposes should at least be stated on a non-disease-specific outcome level (see section 3.4) and be related to the patient categories to which the technology applies[6]. Purposes may thus be translated into measurable entities to be used in the subsequent quantitative analysis of this issue (flowchart analysis, see below) and outcome analysis.

From the perspective of an MTA the description of the selection criteria and the selection process within the health care programme is of critical importance. The results of all other issues depend on the selection criteria and their application.

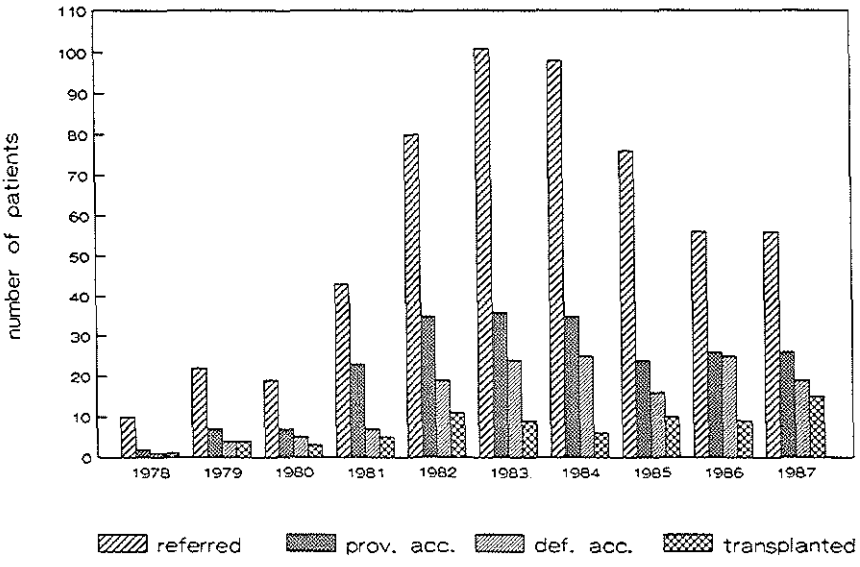
A systematic approach of these four topics often results in a broader description of the technology than might be expected at the onset. This is particularly caused by the attention paid to the alternative modes of action if the new technology is not or only partially applicable[7]. In practice we usually arrive at a description of a comprehensive health care programme for a particular patient group (for examples in the case of liver transplantation, see Health Council, 1989, *A1*, and Bonsel, 1988E)[8].

The quantitative part of the description of a technology primarily consists of flowchart analysis. Such analysis requires a simplified presentation of the patients career within the health-care programme related to his/her condition and prospects. A quantitative model (see section 3.2) of the health-care programme should be defined, representing all the essential relations between subjects over time. Good examples are the detailed models used in MTA's of screening for breast cancer and cervical cancer, of therapy for end-stage renal disease, of heart transplantation and of dental care (Van Oortmarssen, 1990; Habbema, 1988A; De Charro, 1989B; Van Hout, 1990; Truin, 1982).
A simpler model was used for liver transplantation (Habbema, 1988B).

In the model the patient's career should start at the point when the new technology provokes change in routine medical practice. A model which starts at the stage the new technology is actually applied, usually leads to an underestimation of its impact. This is obvious in transplantation MTA's where pre-transplantation activities place a visible burden on the programme, but this is less clear in many other programmes.

Patient flow (transition probabilities related to time) provides information on selection criteria, waiting times, and ultimately programme costs and programme outcome. It may also disclose changes over time and differences between providers. The transplantation MTA's give numerous examples (see Figure 3.1 and Van Hout, 1990).

Figure 3.1 Annual inflow of patients in the liver transplantation programme, according to stage of selection. 1978-1987.



Technology - dependent characteristics

For preventive and diagnostic technologies, health care programme description will be more complex than for therapeutic and rehabilitative technologies. For preventive and diagnostic technologies selection criteria are usually ill-defined (indeed they often have yet to be established) and the paths leading to the ultimate outcome are not easily described.

3.4 Effects on patient outcome

Relevance

Patient outcome is the most important issue in MTA. If the effectiveness of a new technology is beyond doubt and no alternatives are available, a (conditional) positive societal decision on that particular technology is likely. The reverse is also true: without improved patient outcome compared to existing alternatives, analysis of other issues is justified only in exceptional situations[9].

In this section we shall give an introduction to two issues in the methodology of patient-outcome analysis: the concepts of outcome and the choice of a suitable design. An elaboration of these issues is found in the chapters 4 and 5.

Methods - concepts

All efforts of those working in the health care system are directed towards the improvement of the health of the population, be it directly or indirectly. Few would debate this general statement. Difficulties arise if

we try to translate the intuitively appealing notion of 'health improvement' into the language of science. The easy notion in everyday language then turns into a concept escaping unambiguous formal definition.
Our approach to the problem is to start with the scientific context of its use (in MTA) in order to arrive at two different concepts: a concept from a medical and an economic point of view[10]. Both concepts have their own merits and should be used in MTA.

From a medical point of view patient outcome can be broken down into survival, quality of life/general health status (both are disease-non-specific concepts), and disease-specific health status[11]. Except for quality of life/general health status, these concepts seem unambiguous.
Besides, in medical research a wealth of biomedical process parameters are in use[12] e.g. blood chemistry, coagulation status, blood pressure in the portal vein in chronic cirrhosis. Depending on the disease under investigation, the medical researcher chooses one or more of these parameters. Generally, the view has gradually been accepted that the outcome on the patient level should be decisive in evaluative clinical research. Consequently a shift from the use of clinical process parameters towards the use of patient-oriented outcome parameters in the clinical evaluation of medical technologies may be observed. This pertains particularly to the use of parameters of survival and disease-specific health status. For obvious reasons, this change applies less to the use of parameters of non-disease-specific health status.

Within the medical concept of outcome in MTA, survival occupies about the same position as in clinical research but non-survival aspects are treated differently. The emphasis in MTA is on measurement of disease-non-specific health status: those who decide on medical technologies are not only interested in the disease-specific value of a medical technology (compared with its best alternative), but also in its general

benefit (compared with other technologies for other purposes). From the clinical decision-perspective the use of a disease-specific parameter often will suffice. However, the second perspective causes the emphasis in outcome analysis (based on the medical concept) on standardized, generally applicable, quantitative parameters rather than local, disease-specific parameters[13].

If an MTA includes an economic evaluation, an economic concept of outcome should additionally be developed, which should fit into the frame of economic measures for desirability or preference.

The natural measure of economic desirability of a good ('commodity') is its price. In a well-functioning market, economic theory assumes the price of commodities to reflect their actual value and the dynamics of demand and supply will set the price at precisely this value. By comparing prices, we establish the relative desirability of commodities. However, the health-care system does not fit well to the conventional notion of a market (this fact contributes to the existence of MTA). Supplemental economic theory is needed to describe and measure the economic value (economists often speak of 'utility') of health-care services, as their price is not a good indicator (Mooney, 1986 [p.21-33, 64-71]).

By far the most important economic framework to judge the desirability of the commodities associated with health care (including new technologies) have been adaptations of so-called cost-benefit analysis (CBA, see Mishan, 1971 for an account). In its usual form, this framework requires the analyst to arrive at monetary units as the one parameter of benefits. Keeping the structure of CBA grossly unaltered, health economists have accepted other parameters which all are thought to represent economic value or utility. The three accepted *general* parameters are life years gained, quality adjusted life years gained and the healthy years equivalents gained[14].

If life years gained are used, we speak of cost-effectiveness analysis (CEA), if quality adjusted life years gained or healthy years equivalents gained the term cost-utility analysis (CUA) is preferred. With life years alone an extension of standard survival analysis will suffice to describe outcome within the economic concept. If not, complex procedures are needed (see chapter 5).

So far health economists have been reluctant to apply cost-benefit analysis (CBA) in its traditional form, but if we use past decisions on health-care and non-health-care investments as *revealed preferences*, the value of life indeed may be estimated and subsequently used.

Methods - design

An experimental design is the preferred design of measuring net patient-outcome differences, be they medically or economically defined. However, at least two conditions must be met if data are to be suitable. First, the control group should receive the best alternative treatment. Second, effectiveness should be measured in the long run. These objectives often are difficult to achieve. Chapters 4 and 5 deal with technical solutions to overcome the problems if the design is not optimal in this respect. The individual predictability of the course of disease will appear to be the primary determinant of success.

Technology - dependent characteristics

The empirical approach to outcome analysis depends on the technology involved. With a therapeutic or rehabilitating technology, outcome may be easily defined, regardless of the concepts applied.
If the object is a preventive, or diagnostic or compound technology (e.g.

screening), measurement at the outcome level is more difficult due to the time which has to pass before effects may be expected. The use of intermediate parameters (for example process indicators) may then be indicated.

3.5 Effects on the use of resources

Relevance

This issue forms the counterpoint to patient outcome analysis in MTA. Though some doctors still hold the opinion that cost considerations are not ethical, the economist's view has gradually gained support. Resources are limited; any decision causes changes in resource allocation; ignorance of these changes may lead to a larger disadvantage for some unspecified group compared with the advantage for the patient group of the new technology; hence the *absence* of cost considerations may be called unethical. In addition to this ethical argument it may safely be stated that cost considerations play a major role in most doctors' decision making. Finally, the inclusion of the issue is justified by the level of total costs of health care. Modern societies devote a large proportion of their resources to the production and distribution of health care.

Though cost considerations may sometimes be second to effectiveness considerations, they form a necessary part of MTA.

Methods - definitions

From an economic point of view, cost-analysis should not be restricted to the actual resources employed by a new technology. Moreover, eco-

nomic theory provides an analytical framework to support the decision-maker, which constructs parameters for optimal resource allocation. This framework needs information on input (costs), throughput (process of application) and output (consequences, e.g. patient's outcome).
Following Drummond, such an analysis is called *full economic analysis* (or economic evaluation; Drummond, 1987A [p.8] & 1987B; De Charro, 1989B). Partial economic analysis is restricted to the input or output alone, it takes only the new technology into account, or it applies both restrictions. Partial cost-analysis does not yield a criterion of choice, hence full economic analysis should be regarded the standard approach. Full economic analysis may be structured as cost-effectiveness analysis, cost-utility analysis and cost-benefit analysis, depending on criterion of outcome and the purpose of optimization.

Within an economic evaluation the determination of economic costs should be distinguished from an account of the financial costs.
Economic costs refer to the real input of all resources valued according to the opportunity cost principle: what could have been done with the same resources in the best alternative way? A perfect market more or less guarantees actual prices of the resource input to reflect economic costs. The health-care system is not organised as a perfect market due to among other things monopolies and the agency function of doctors (see e.g. Mooney, 1986 [p.83-87]; McGuire, 1988 [p.31-53], Williams, 1978). This implies that economic costs in health care have to be distinguished from the financial costs[15]. Financial costs should preferably be estimated too. The decision-maker who commissioned the MTA (but other parties too) may need these estimates for other purposes.

Methods - design

Several authors have listed the elements of the empirical part of economical analyis (Drummond, 1987A [p.22,39-]; Rutten, 1989 [p.105-]; Luce, 1990 [p.65-67]). Once the values of all the elements have been established, results are aggregated. Depending on the outcome criterion chosen and the purpose of optimization, we arrive at a particular summary measure (e.g. a cost-effectiveness ratio). As an important economic feature of aggregation allowance is made for different timing of events[16]. In this thesis we will not discuss further the empirical procedures associated with the monetary part of the table: generally the ten guidelines presented by Drummond form suitable anchorpoints (Drummond, 1987A [p.35-38]). The non-monetary part of economic evaluation is further discussed in section 3.4 and chapters 4 and 5. Aggregation, for example the computation of cost-effectiveness ratios, will be given some attention in section 3.8.1. The remainder of this section is devoted to some conceptual remarks from the viewpoint of a medical professional.

Costs of care vs. costs of research; scale effects

With new technologies it may be difficult to separate resources invested in patient care and resources invested in clinical research. In as far as final decisions frequently depend on costs considerations, the distinction is important. Neglect may lead to the unjustified rejection of a programme for its assumed expense or, with the acceptation, to large scale subsidies for clinical research - which are not at issue in *this* decision-cycle. The more immature the technology, the harder it is to make the distinction[17]: detailed economic analysis is not very useful if technologies are still developing rapidly.

This part of economic analysis requires medical expertise, which should prefer be provided by the experts themselves. However, some reluctance to cooperate may be expected. Even if the circumstances are advantageous, it may be a difficult empirical task.

The issue of scale effects resembles the issue of distinction of research costs from standard care costs. Scale effects are likely if care is, or is to, standardized and if care is, or is to, provided on a centralized basis. The estimation of the potential scale effects requires analysis of the functioning of the health care programme over time or in different settings, supported by information from medical experts. The latter may be reluctant to cooperate if they expect their academic and clinical freedom to be violated by conclusions drawn from the results. This potential conflict of interests may explain the absence of these issues in published MTA's. In some cases international comparison of similar programmes in different countries may provide clues to scale effects (Bonsel, 1988A).

Future medical costs

This issue may be best introduced by an example. Imagine a perfect life-saving technology which prevents preterm delivery with fatal outcome. As a result a healthy child is born. From an economic point of view the normally expected life-time medical costs of the newborn should be regarded as direct medical costs to be included in the analysis. At first sight this inclusion seems just an odd feature of health economics.

The similarity with an analoguous feature of public health theory is remarkable. In the context of judgment of the effectiveness of medical care at the public health level it has been suggested that the number of diseased and disabled will rise due to the effectiveness of modern medicine (the issue is described by Van der Maas, 1989). If this is true

we might be tempted to disregard the lives of miserable quality saved to avoid the conclusion that the balance of sophisticated medical care is negative.

Both paradoxes require carefully devised aggregate measures[18]. The inclusion of postponed negative health effects and medical costs may thus be accounted for (Roos, 1990).

Technology - dependent characteristics

As in outcome analysis, the type of technology influences the empirical task in cost-analysis (see section 3.4). Due to the multiple ramifications of the chain of cost-incurring activities, the empirical analysis of preventive, diagnostic and compound technologies may be demanding.

3.6 Need for the technology

Relevance

Determination of the need for a particular technology may serve many goals.

First it allows for an evaluation of the actual performance of the clinical pilot programme with regard to its regional or national coverage (see also section 3.3). Secondly it allows for future projections of patient flow, provided a forecasting model is available (see section 3.2). Consequently the magnitude of resource use and of the consequences may be estimated. Four examples show how need estimates may be relevant to the decision-maker.

If an analysis of need suggests impact on the population as a whole, as with preventive technologies or technologies for prevalent diseases, impact considerations may be recognized as a separate issue. If need projections suggest that particular constraints will inevitably limit the access to the technology, distributional considerations may be regarded as an important issue. If the projected size is small, centralization of technology application may be considered, or, with short term interventions, treatment abroad. If need projections suggest a rapid growth, organizational arrangements may be considered to guarantee equity and efficiency.

These examples suggest that from society's point of view need estimates are essential for considerations of efficiency and equity. From a supplier's point of view, need estimates may serve similar goals like market development and organizational optimization.

Below an outline of the two main concepts of need is presented. As will be explained in chapter 6, general remarks on the empirical design are not relevant in this case.

Methods - concepts

As with outcome analysis we distinguish a medical and an economic concept of need (see also chapter 6).

From the medical point of view we look for individuals who share characteristics which define them as suitable for the application of a particular technology; i.e. a significant gain of health benefit may be expected from the application of that particular technology. This concept is rather absolute as it excludes other considerations than those related to direct patient benefits. Obviously, some extremely effective medical services do not need these additional considerations.

From the economic point of view the concept of need is a theoretical supplement to existing classical theory on demand, supply and functioning of the market (Williams, 1978; McGuire, 1988 [p.150-166]; Mooney, 1986 [p.72-87]). Need in this view has to do with expected health benefits *relative* to the resources to be employed and to existing alternatives.

Technology - dependent characteristics

An empirical analysis of need of preventive and diagnostic technologies frequently implies a separate epidemiological study, even if the medical concept is applied. Empirical analysis of need in other technologies is usually less complex, but data are not abundant (see chapter 6). Lack of data is also the major limitation to the use of the economic concept.

3.7 Current and future uncertainties

Relevance

In this section attention will be paid to three related topics: sensitivity analysis, trend analysis and policy analysis. They all deal with uncertainty. Uncertainty plays a major role in health care decisions and its general disutility in the decision process may easily be observed (O'Brien, 1988)[19]. The quantification of uncertainties is an essential means to reduce this disutility (Drummond, 1987A [e.g. p.7]). Although in a technical sense these uncertainties are frequently treated in a similar way, their significance differs.

Definition of sensitivity analysis, trend analysis and policy analysis

Sensitivity analysis examines the actual uncertainty of the values of key parameters which are produced in the analyses described in sections 3.3 to 3.6. It may be distinguished into three types.

Uncertainty about the relation between the distribution of observed data and the distribution of the values in the real world (measurement errors).
We need to know for example what a five year survival rate of 69%, based on partially censored survival data of 73 patients, tells us about the certainty that a newly treated similar patient will survive the same period. Standard statistical approaches, including, meta-analysis are usually available to address this type of uncertainty.

Uncertainty about the distribution of the values in the real world because no measurement instrument is available ('intangibles').
Intangibles refer to difficulty in counting, valuing or both. For example temporary health effects, as encountered in a vaccination or a screening programme, are difficult to count. Process effects are intangibles in another sense as they are countable but difficult to value.
Intangibles may be treated in several ways (Drummond, 1987A [p.153-156]). They may be omitted in toto, they may be measured indirectly and they may be estimated by a panel.
If all alternatives share the same intangible, neglecting it may be justified. Indirect estimation of an intangible may sometimes be a valid alternative. For instance patient's behaviour may reveal the disutility of traveleing time, the disutility of waiting for treatment given an opportunity to receive treatment elsewhere, etc. Frequently these approaches do not satisfy. 'Guestimation' will then be the strategy. Eliciting values from experts may be useful, particularly if formal techniques are applied (Lammerts, 1987). Scale and learning effects may be estimated this way.

Uncertainty about the absence or presence of values in the real world due to dissensus about their existence and/or the relevance of their inclusion in the MTA ('imaginables').

Imaginables may be found in outcome analysis (quality-of-life-effects of relatives and friends, life years gained of a future generation, e.g. with in-vitro-fertilization, particular process utilities), cost analysis (production losses) and analysis of need (e.g. abortus provocatus).

Their treatment rests upon convention. These conventions should preferably be the result of discussion between MTA-investigators, experts in the object field and the decision-maker. A reluctance to this undertaking may be observed: the MTA-investigator defines the issue as political, the decision-maker as scientific and the clinical expert wonders what the problem is at all.

Uncertainty about what the real world will look like in forthcoming years is another type of uncertainty. Forecasting techniques may be helpful to the decision-maker in this case.

Forecasting the effects of autonomous developments is called *trend analysis*. The technical elaboration is frequently based on the projection of the consequences of demographic change only, assuming a stable relation between the population composition and the technology involved[20].

If, additionally, the effects of policy decisions are included, we speak of *policy analysis*. Some options to be included are listed in Table 3.4. Many health care decisions are made with policy options in mind. Interplay between MTA-parties will make these options more explicit and in the favourable case, the effects of these options are estimated a priori.

Table 3.4 A selection of policy options to be included in MTA

Regulatory activities towards technology
- price/volume/conditions
- patient/doctor/provider

Health promoting activities
- technology related/unrelated
- risk reduction/risky behaviour reduction
- individual/collectivity
- optimization of demand/supply

Activities on the demand for health care (e.g. insurance arrangements)
Activities on the supply of health care (e.g. man power planning)

Educational and socio-cultural activities

Methods of sensitivity analysis, trend analysis and policy analysis

The basic tool is the quantitative model. Variation of the structure, the relations and the data values may mimic real world changes. Micro- and/or macrosimulation may be used to compute estimated values in the future. The type of the model (stochastic, deterministic, mixed) determines how uncertainty is exactly accounted for. The particular choice is based on practice, not on in principle[21].

The combined prospective results on all issues within a set of assumptions is called a *scenario*. The *reference* scenario combines existing technological alternative(s) with trend analysis. The *basic* or *null* scen-

ario the introduction of the technology combined with trend analysis.

Threshold analysis may be undertaken within the context of sensitivity analysis but also within the context of trend and policy analysis. It presumes a model in which for one or more key parameters - particularly aggregate measures of outcome, etc. - values are pre-specified below or above which the decision taken will be different.

In fact these key parameters are the operationalization of criteria formulated in Phase 2 of the decision cycle. In our experience these criteria are only infrequently formulated. Consequently threshold analysis (like marginal analysis[22]) still has limited applicational value in MTA's[23]. The methodology is described elsewhere (e.g. Pauker, 1987)[24].

A general feature of the analysis of current and future uncertainties is the difficulty in achieving a balance between scientific credibility and political suitability, which may be a Procrustean exercise. The different approval of the final MTA-reports on the Dutch Heart Transplantation Programme and the Dutch Liver Transplantation Programme provide us with an interesting example.

Both reports were conceived and written in a similar fashion by essentially the same research group. However, they differed in how the uncertainties were presented.

The heart transplantation report presented a point estimate of the cost-effectiveness ratio of 52,000 Hfl/life year gained. This central estimate was the average result of 100 microsimulations, substituting values and distributions of values in the model with, according to the authors, the highest likelihood of occurrence. The potential consequences of all types of uncertainty was computed by carrying out additional simulations, each time changing one (or more) distribution of key parameters[25].

The liver transplantation report did not apply stochastic modelling techniques and presented a range of the cost-effectiveness ratios of

50,000 - 130,000 Hfl/life year gained. The lower range all favourable ones, the upper range combined all unfavourable assumptions.
The point estimate of the heart transplantation report did survive as an exact result as opposed to the 'inexact' range results in the liver transplantation report. The apparent exactness in the first case was mistaken for certainty. It contributed to a perceived superiority of the first approach, both in a scientific sense (certainty as parameter of scientific skills) and a decision sense (certainty as parameter of suitability). We regret the argument, not the conclusion.

3.8 Miscellaneous issues

In the final section of this chapter four issues will be addressed: formal synthesis, subgroup analysis, analysis of distributional effects, and social, legal and ethical analysis. The first and second subsection address two statistical tools which receive special attention in MTA. The third and fourth subsection deal with two optional issues (see section 3.1).

3.8.1 Formal synthesis

As MTA provides us with information on many topics, the synthesis of results from *different issues* evidently would be useful to the decision-maker. Several aggregational techniques exist to deal with the problem: multicriteria analysis, cost-effectiveness analysis, cost-utility analysis, cost-benefit analysis (see IOM, 1985).
Within the framework of economic evaluation, all but the first technique may be regarded as standard approaches to integration. Although the calculation of these summary measures have been criticized, it should be

stressed that aggregation is a basic scientific tool which is also applied throughout the analysis of separate issues.
This subsection starts with a short account on aggregation within the issues, followed by an account of 'multi-issue' aggregation.

Aggregation of results, for example by averaging values and computing confidence intervals, may be regarded routine practice in conventional methodology. Hence, acceptance of aggregational operations in the separate analyses in MTA is obvious. Two exceptions deserve attention.

First, multidimensional health status concepts have been rejected either explicitly or implicitly by redefining the issue of overall judgment on health status as an ethical issue (Health Council, 1989 [p.141-]). However, multidimensional concepts are ubiquitous in science, medical science included; the new thing is statistics and computers providing practical *means* to deal with these concepts rather easily (chapter 5).
Secondly, counting up health status measures of different persons has been rejected (Harris, 1987). Again we regard this position to be untenable by the argument that the construction of a survival curve and of all other irreproachable aggregate measures of health consequences are conceptually no different from aggegrated health status values.

A second type of aggregation is the combination of (aggregate) results of different issues. This type of aggregation is encountered in formal synthesis of MTA-results. The aggregation techniques mentioned before all are quantitative methods of data-reduction, yielding an overall value or judgment.

Multicriteria analysis is based on a prespecified weighting of outcomes on an essentially arbitrary set of important issues; operations research provides us with various optimization techniques. We are not aware of explicit application in health care[26].

Cost-effectiveness analysis (CEA), cost-utility analysis (CUA), and cost-benefit analysis (CBA) are all based on the foundations of welfare economics (Drummond, 1987A [p.150-]): their attention is restricted to all countable inputs and outputs, relating (usually by division) the net use of properly valued resources to net outcome. Essentially the acceptation of CE-, CU-, and CB-ratios is based on the assumption of inputs and outputs as representation of ends and means and of division (linear proportionality) as the most reasonable function rule.

These reductionistic quantifications of the complexity of a health care programme have been particularly prone to criticism (e.g. during the transplantation studies, see also Van Rossum, 1989).

Part of the criticism on CE-, CU-, and CB-ratio's probably originates from a categorical mistake on aggregation in general, exchanging the real world and its abstraction[27].

Another part perhaps arises from misperception of the political significance of measures economic desirability. A low cost-effectiveness ratio does not *necessarily* mean the apparently most efficient technology is the 'best' and consequently should be chosen (e.g. Keeler, 1987). Caution with these ratios is particularly advocated if health-care programmes with different targets are compared (see e.g. Leaf, 1989).

In a technical sense the current lack of cost-effectiveness ratios prevents us from justifying the latter type of choices on the few available data[28]. In a political sense we must accept that other, often implicit, criteria may depreciate what seems to be the optimum from cost-effectiveness considerations alone[29].

In this regard we would cite McGuire, who concisely describes the boundary between MTA-analysis and decision-making (1988 [p.99-100]).

"The decision-makers responsible for the final choice of project(s) may not, of course, choose the project(s) declared to be efficient. The data

on which the analysis was based may be inaccurate and the decision-makers may feel that, with their more extensive knowledge, some other project is more efficient. Or they may disagree with the value judgments that the analyst will necessarily have introduced into the appraisal. (...) Or they may decide that a more equitable, but less efficient, project should be chosen. Thus the claim of CBA, CUA, and CEA is not that they *make* decisions, rather it is that they assist in the decision-making process. Perhaps the chief merits (which should not be undervalued!) are that they make health care planning comprehensive rather than partial, systematic rather than piecemeal, and value judgments explicit rather than implicit."

In summary it is concluded that aggregation is an indispensible basic tool in any scientific activity. Its use in MTA is indicated at the issue-level and is advocated at the level of formal synthesis.

3.8.2 Subgroup analysis

This subsection treats a technical feature of MTA which to some extent mirrors the feature discussed above: subgroup analysis.

As a rule, medical research on the effectiveness of an intervention, tries to answer two questions "Does it work?" and "For whom does it work?". The scope of conclusions of MTA is broader, but they apply the same division.

First we are interested in overall performance of a health care programme; formal aggregation techniques as discussed in section 3.8.1. may be used to arrive at an answer to the question "Does it work with a reasonable employment of resources?"

Next we are interested in the question on local optima: under which conditions (of any kind) performs the programme best?

The issue is referred to as subgroup analysis; both statistical and political considerations appear to be taken into account. We start with some statistical considerations.

Experiments may be designed *a priori* to allow for testing the effects of selected variables. Usually selection is *a posteriori*. Analysis a posteriori takes with it the danger of finding 'significant' variables due to chance alone, particularly when an impressive number of unknown confounders is tested in a less impressive number of patients, as inevitably is often the case in MTA. Regardless of the issue involved, prior specification of these confounders at the onset of the MTA is preferable[30].

A priori specification may also be justified to avoid a posteriori objections from e.g. the health care provider or society's representatives if results are expected to be in conflict with their interests. Here we arrive at political considerations. One *fictitious* example may illustrate the point. An additional one is contained in the note[31].

A first explanatory regression analysis of the waiting time for a donor liver shows that women wait longer than men. Also, the older people wait considerably longer. Entering diagnosis in the explanatory model appears to decrease the influence of sex and age below significance, with primary biliary cirrhosis taking the role of primary explanatory variable. When, in the third stage of analysis, severity of disease is added as the fourth predictive variable, it removes diagnosis from the list of significant variables, and age reappears as a second significant variable after severity of disease. Different moral reactions to these three stages of analysis may be imagined.

In our research activities we also encountered these non-statistical arguments against the use of a particular explanatory variables[32].

We conclude that subgroup analysis may be justified from a statistical point of view, particularly if empirical data on key parameters show

much variation which is accounted for by a priori specified variables. Non-statistical arguments may limit the degree of freedom in choosing the explanatory variables.

3.8.3 Distributional effects

In this subsection the analysis of distributional effects is briefly discussed. We start with an explanation of the relevance of the issue which is also known as the equity-issue. Next equity is defined and some remarks on the effects of equity considerations in MTA-context are given. Finally we reply to the presumed inequity consequences of the use of QALY's.

In chapter 2 the prior definition of core issues and the criteria to be used in decision-making was stressed. If we accept more than one criterion, it will sometimes be difficult to satisfy multiple criteria simultaneously.
Suppose we have chosen as a criterion-variable that average investment needed per QALY should not exceed Hfl 50,000. Suppose we also adhere to a second criterion: health-care programmes directed towards rehabilitative care receive priority treatment. It may be expected that the complete satisfaction of both criteria simultaneously is impossible.

More generally stated, health-care programme selection according to any primary criterion as meant in chapter 2, will usually result in suboptimal satisfaction of secondary criteria. The point is raised for the assumed conflict of two criteria regarded to be important in health care decisions, efficiency and equity.

Equity is also be discussed because the use of a QALY as a unit of outcome has been criticized for the presumed deleterious consequences towards equity. The selection of technologies based on the cost per

QALY criterion would unduly favour the young with acute disease; the aged and those with chronic disease would suffer the use of this efficiency criterion (e.g. Musschenga, 1987).

The rationale of an efficiency criterion is in maximizing society's benefit as a whole, given limited resources[33]. The use of an efficiency criterion originates from utilitarianism. The efficiency of a health-care programme is measured by economic evaluation and may be described by the afore-mentioned ratios (see section 3.8.1).

A definition of equity in health care is less unequivocal (Wagstaff, 1989; McGuire, 1988 [p.63-73]). Available definitions usually[34] deal with the degree to which the distribution of health-care related goods (here the technology or its alternative) satisfies the following two rules:

a. patients who are unequal according to a criterion (usually: need) are treated unequally or, more plainly, the more someone needs health care the more he should get it (vertical equity);
b. according to remaining criteria (particularly: income, gender, age, race) patients are treated equally (horizontal equity).

Apart from the commodity to be distributed and the criteria to be applied, a function rule should be defined which states *how much* more health commodity you get the more you need it[35].
Graphical and computational techniques exist to measure the degree of undesirable inequity.

The following six observations (see next page) in our view suggest that routine application of distributional analysis *within* MTA is not justified yet. The first two observations refer to the concept of equity, the latter to the relation of equity considerations with the standard effectiveness and efficiency considerations in MTA.

1. No standard definition of the vertical equity criterion 'need' has been accepted. The literature provides us with various alternatives: e.g. actual health status[36], health lost without intervention, potential health to be gained with intervention (Mooney, 1986 [p.108]; McGuire (1988 [p.123]). The lack of standards for horizontal criteria is less urgent, as age, gender, race (see for transplantation: Kasiske, 1991) and income appear to be the most important.
2. Horizontal equity criteria may correlate with the vertical equity criterion. E.g. if age determines health lost without intervention or potential health to be gained with intervention (all true in liver transplantation), perfect distribution according to need[37] implies that the satisfaction of the horizontal criterion is compromised. This correlation will be encountered often (also with gender and income), hence the overall degree of equity is apparently frequently not defined.
3. In MTA alternative modalities to solve the same medical problem are compared. Conclusions about the best choice based on efficiency alone may change if vertical equity is taken into consideration. We expect this to be the case only infrequently: particularly if vertical equity is defined as "the more potential health may be gained, the more care you get", conclusions remain unaltered.
4. In MTA efficiency will be in conflict with horizontal equity considerations if the effects of alternatives differ with regard to their relation to the particular horizontal criteria. As this will be often the case, a similar problem as mentioned in 2. arises[38].
5. Discussions about equity seem to focus on the use of MTA-results in a broader context: comparison between different health care programmes in view of resource allocation on the macro level. We agree that equity considerations with regard to age and other criteria for horizontal equity may be relevant in that case (see e.g. Callahan,

1990; Levinsky, 1990). In the end the dominant philosophical position will determine the primacy of either efficiency, effectiveness, vertical equity or horizontal equity. However, if horizontal equity is accepted as the primary criterion (to be technically achieved by standardizing health status or outcome measures according to horizontal criteria), medical science would be redefined[39].

6. The afore-mentioned problems arise with any criterion for need, be it health-status or health-improvement related, hence the selection of particularly QALY's as measure of outcome is not essential to the issue.

3.8.4 Legal and ethical issues

The position of legal and ethical issues in MTA-reports is not as clear as the previous issues (e.g. Luce, 1990 ignores these issues).

We think that *legal* issues which are relevant within in an MTA should be distinguished from the legal features relevant to the decision process which an MTA is supposed to support. For example, within the transplantation studies we encountered the issues of brain death criteria (Health Council, 1983; id., 1987[40]) and of legal aspects of acquiring and distributing donor organs (Health Council, 1987)[41]. However, in many cases technology-specific legal issues are not at issue in an MTA.

If an empirical MTA is carried out, legal features of medical investigation *in general* deserve attention, e.g. the use of registration data, and of patient data (Vandenbroucke, 1989).

Numerous legal issues can be considered *before* and *after* this information phase of the decision cycle. Their relevance to the consecutive phases of the decision-cycle is described in a note[42]. If selection of patients is involved, the legal analysis may be complex[43].

Judging from the contributions of legal and ethical experts, legal and ethical issues within an MTA frequently coincide and their approaches has much in common (compare e.g. Roscam Abbing, 1989 with De Wachter, 1989). Country-specific characteristics of the legal system may limit the use of foreign legal analyses (compare e.g. in heart transplantation Evans, 1984 with Health Council, 1988).

Ethical considerations occupy a special position.
Moral judgment has been at the core of philosophy for centuries. Medical ethics have grown only recently to maturity. Its development is closely related to the development of professional medical organization and the medical knowledge base (Ten Have, 1988 [p.24]). Traditional deontological professional ethics are embodied in the famous Hippocratic oath. At present medical ethics have extended from a science on normative theories in medicine to an applied science on the process of application of these theories (Ten Have, 1988; Beauchamp, 1989; De Beaufort, 1989). The former has been called the ethical product, the latter the ethical process.

The IOM lists at least three theoretical principles of medical ethics as a deontology (the ethical product): that of autonomy, justice (connected with the issue of equity) and the pursuit of common good (connected with the issue of individual and societal benefit) (IOM, 1985 [p.157]). The relevance of analysis of these principles in an MTA will vary with the technology involved. Informative technologies, procreative technologies, and technologies to be applied in children and psychiatrically ill for example are particularly relevant to the autonomy principle. Ethical analysis may then be indicated. The relevance of the equity issue as a subject for ethical analysis is beyond doubt but its analysis is not without difficulty (see 3.8.3).

The *ethical process* concept has been described by Ten Have (1988)[44]. The concept focuses on *how* we can arrive at an optimal ethical judgment instead of *which rules* determine whether decisions or conduct are ethically justified. The author states two important features of the process concept. Moral reasoning depends on the moral rules, which in turn are derived from moral principles. "Moral reasoning itself provides no criterion for choosing between conflicting principles" (for an example see section 3.8.3 on equity). Moreover, "reasoning offers reasons to people, but people [and: health-care authorities GJB] are not always reasonable". The analysis of Ten Have suggests that the primary importance of medical ethics as a process is in discussing the decision cycle as it stands: as with the legal position, its primary value appears to be meta-analytical.

Here we do not give a full analysis of these principles and their application. Mooney, an economist, concludes that medical ethics sofar have failed to address scarcity satisfactorily (Mooney, 1986 [p.88-106]). This is a drawback of medical ethics in the context of MTA. Thus it might be too early to advocate the routine participation of a prospectively active ethical expert in MTA (Van Leeuwen, 1989).

3.9 Summary

In this chapter a standard approach for empirical MTA is proposed. It is argued that standardization is of vital importance to MTA: it allows for comparibility of results across technologies, diseases, and populations. Besides it is required for economical analysis and forecasting. The construction of quantitative models which integrate medical and economic information is a logical consequence.

Five core issues are considered to be essential in MTA: description of the action of the technology, the effects on patient outcome, the effects on the use of resources, the need for the technology and finally the actual and future uncertainties. The major part of chapter 3 is devoted to conceptual remarks on these five issues.

In the last section attention is paid to two technical procedures: the formal synthesis of results applying summary measures like an cost-effectiveness ratio and subgroup analysis. We end with a short account of distributional analysis and legal and ethical aspects.

4 ANALYSIS OF SURVIVAL

4.1 Introduction

In this chapter some theoretical and practical aspects of survival-analysis are discussed. From a statistical viewpoint survival analysis is the application of a set of tools to analyze the occurrence of an event in time[1]. From the viewpoint of MTA, survival analysis is the cornerstone of outcome analysis (see section 3.4).

Section 4.2 presents some general remarks on survival analysis in MTA, distinguishing between survival as one of the three medical outcome concepts and survival as an economical outcome concept. Attention is paid to the estimation of survival consequences of an intervention in the absence of any information from randomized controlled clinical trials (RCCT's); this absence may be often observed in new technologies. In section 4.3 some technical features of survival analysis are shown.
The technical approach of a method which predicts control-group survival is presented. This method uses prognostic information from each individual directly preceding the intervention. The predictions may be used both at the individual and at the aggregate level (Christensen, 1987). In the latter case the predicted survival curve, of a group of patients may be compared with the observed survival curve enabling judgments of the effectiveness of the target technology. Also the computation of the number of life years gained (or lost) due to the target technology is possible.

Section 4.5 shows results from the liver transplantation MTA, preceded by some historical notes on survival analysis in liver transplantation (section 4.4). Section 4.6 is a summary.

4.2 Survival analysis in MTA

Two concepts

In section 3.4 a distinction was made between patient outcome from a medical and from an economic perspective.

The first concept includes survival, quality of life/general health status (both are disease-non-specific concepts), and disease-specific health status. Survival analysis in this case provides an answer to the question whether survival with the target technology differs from that of alternatives (in general or after a fixed interval) and which determinants (including intervention) explain observed heterogeneity of survival, using cumulative survival percentages as the conventional unit of outcome. For example it may be stated that the survival percentage with sclerotherapy of oesophageal varices is 80% and 40% after 1 and 5 year follow-up (fictitious numbers) and that it does not differ significantly from survival with conventional therapy.

The second concept uses survived years as the primary parameter of outcome. If quality-of-life adjustment are not included (see chapter 5) survival analysis in this concept follows the same route as above but finally integration of the survival curve is carried out. As a result it may, for example be stated that patients with oesophageal varices live an average 4 years regardless of the therapeutic modality, and that consequently net benefit of sclerotherapy is zero life years gained.

Graphically, the difference between the two concepts is simple: the first concept is directed to the level of the graph over time and statistically to the occurrence of terminal events; the second is directed to the size of the area below the graph.

Shadow survival

One of the commonest and most difficult problems in survival analysis in MTA (with both concepts) is the absence of experimental knowledge on the outcome of alternative options. Too frequently an RCCT is rejected because the new technology is believed to be 'much better' than any rival option. The same argument was thought to be valid in liver and heart transplantation at the stage they were investigated in the Netherlands (1985-1988) (Habbema, 1988B; Bonsel, 1989B; De Charro, 1988). As was shown by Van Hout in the heart transplantation case and in this thesis in the liver transplantation case, differences were not always as large as expected (Van Hout, 1990; *A3*)[2]. These examples show the vulnerability of clinical judgment in the absence of RCCT information.

In Table 4.1 we present five methods which deal with this problem (Bonsel, 1990A). All these methods estimate *shadow survival* i.e. they estimate the probabilities of survival supposing the target technology had not been applied.

Table 4.1 Methods to estimate shadow survival.

1. Formal expert judgment
2. Historical control group
3. Quasi-experimental control group
4. Intervention delay group
5. Prognostic modelling

The characteristics of these methods are discussed below. Examples are drawn from the liver transplantation study in which the application of these methods was considered (Bonsel, 1988F).

Formal expert judgment

Obviously some expert judgment of survival benefit is already manifest in the clinical decision to offer a patient a particular treatment. However, a formal quantitative prediction goes beyond this dichotomous decision.

The strength of expert judgment is its full exploit of human perceptive and integrative capabilities. Clinical judgment can take into account all patient information, including information escaping formal definition such as the change of features like the patient's face, texture, and his morale. Hippocrate's treatise on prognosis already points to their importance (Anonymous, 1983B [p.171]).

The weakness of expert judgment is related to its strength. Human judgment is prone to error, and prone to persist in error. Once a decision towards an intervention is made, all information subconsciously tends to be interpreted as a support of the option chosen[3]. Judgment is also influenced by the social environment. Group judgments tend to converge if personal ties or shared objectives place a disutility on dissensus.

For these reasons clinical judgment by experts of shadow survival can only be used within a formal design. For example: in liver transplantation survival estimation of a sample of anonymous historical controls of a liver unit can be used to discover the prognostic evidence used by doctors.

A Delphi-technique (see section 2.6) may subsequently be used to arrive at group estimations and would-be decisions on these patients.

Historical control group

Survival information from similar patients treated in the past (reference patients) may sometimes appear to be suitable for direct approximation

of shadow survival in current patients (index patients). Nevertheless the representativeness of both the patients and their prospects should be evaluated. Table 4.2 lists some sources of decreased representativeness.

Table 4.2 Sources of bias using historical control group data.

Patient	*Prospect*
Patient selection	Changes in epidemiology
Diagnostic changes	Changes in therapy

A critical problem is how to evaluate the selection of patient data. As a rule patient data are not derived from complete registries. We also do not know, whether patients presenting their disease or problem *now*, do represent the patients in the *past*.

Secondly, diagnostic criteria may change. New technologies may induce diagnostic refinements. In this area for example, the development of the endoscope contributed to the disentangling of primary sclerosing cholangitis, recurrent infectious cholangitis and cholangiocarcinoma.
Even if criteria are left unchanged the diagnostic and prognostic profile may change: refined techniques to discover peritoneal metastases in ovarian cancer yields 'improved' survival rates in all stages of the disease, without any effect on overall mortality (De Kroon, 1990).

Epidemiological change is relevant, particularly in infectious diseases[4]. Even in non-infectious diseases patterns of disease may change (Van der Maas, 1988) invalidating the use of old data on prognosis[5].

A common bias in the use of historical data is the underestimation of the effectiveness of historical therapy at present. Gradual improvement of any therapy may be expected, even if lethality has not decreased. For

example, the improvement of results in the medical treatment of cirrhosis and in the palliative treatment of liver carcinoma invalidate the straight forward use of historical data (as in De Jong, 1990).

The strength of historical data is their empirical nature. Their reliability and validity can be estimated. Frequently time to follow-up suffices to gain a fair impression of average life expectancy with the disease.

Quasi-experimental control group

The limited application of new technologies may offer the opportunity to compare the survival of the non-recipients of the new target technology with those receiving it.

If the benefit of a new technology is unclear and public appeal is low, doctors not involved in the clinical pilot programme will be reluctant to refer patients. If additionally there is a clear competing option outside the clinical pilot programme and if patient selection is identical, quasi-experimental data may be validly derived. We made use of this method in biliary atresia (see section 4.4).

Another source of shadow survival data are patients who are referred to the intervention programme but who do not receive the target technology due to withdrawal or deselection. In the case of heart transplantation accepted candidates, who decided to reject the opportunity for this treatment, provided valuable information on shadow survival (Stevenson, 1986). The same holds for patients rejected for reasons *only* unrelated to survival with transplantation (e.g. psychiatric disorders). Even patients, with a prognosis without intervention considered too well to justify intervention, may provide important information on the upper limit of shadow survival (Stevenson, 1987). In our study the opportunity to use these options was limited (see section 4.4).

Intervention delay group

Waiting periods due to scarcity may be another source of information on shadow survival. Particularly in life-saving technologies, much can be learned from the waiting-time experience of patients, providing two conditions are met[6].

First, patients should be homogeneous with regard to prognosis until intervention when they enter the waiting-list. Their entry should *only* reflect suitability to undergo the intervention, be it after minutes or - without intercurrent change - after months. Particularly if entry selection occurs for some additional reason, this generally results in unacceptable heterogeneity with regard to survival on the waiting list.

Second, deselection due to the intervention (the primary censoring event) or other reasons (withdrawal, or change of the clinical condition) should be *aselect* with regard to the prognosis without intervention.

If these homogeneity assumptions are met[7], several technical procedures exist, both parametrical and non-parametrical, which enable the use of this waiting-list experience (Turnbull, 1974). Some of them were used in the liver transplantation study (see section 4.4 and *A1, A3*).

Prognostic modelling

Prognostic modelling is a method which may be regarded as a refinement of the three previous methods. Potentially it is the most powerful method to estimate control group survival (see e.g. Christensen, 1988).

The method starts with the application of Cox' proportional hazards model (see section 4.3) to the survival times of a group of subjects receiving control treatment in the recent past (for a reference group, see the three previous methods). From this analysis a set of prognostic fac-

tors (patient variables) is derived with their associated weights (regression coefficients), and a basal hazard function describing survival in absence of these prognostic factors[8].

Next the shadow-survival of each index patient is estimated. The actual levels of the prognostic factors of that particular index patient are substituted in the prognostic model derived from the reference patients, and an estimated survival curve is computed. Thus we use historical survival information, controlling for *all significant prognostic factors* distinguished in reference patients.

Three conditions have to be met before this method provides supplemental information as compared to the previous methods.
First, a prognostic model is relevant. If survival of a particular condition is rather uniform, e.g. all patients die after 3 months within a small range, the individual prognosis reduces to the average group prognosis. However, if the survival times distribution of a particular condition does show large variance, this does not guarantee the existence of a prognostic model. Survival may be the result of chance only, it may be the result of unknown, hence unmeasured, factors or a small number of observations may simply preclude the estimation of such a model for technical reasons. If the course of the disease at a particular stage turns out to be *unpredictable* at the individual level, average group prognosis again is the best estimate of individual prognosis. Finally a prognostic model may not exist due to *lack of perceived interest*. In the majority of diseases the only available knowledge consists of crude mortality estimates.

The second and third conditions are those previously mentioned in the crude historical control group method: the diagnosis of index patients should be similar to that of the reference patients and the prospects of index patients should be represented by the prospect of reference patients (apart from the prognostic factors of the model).

The second condition assumes diagnostic equivalence between the patients on which the model is based and the patient to which the model is applied. Etiological heterogeneity is only a problem if it induces survival differences not accounted for by the prognostic factors. This condition also suggests that the whole range of variables in index patients should fall within the range of these variables in reference patients[9,10]. If this condition is not satisfied additional assumptions must be made with respect to the effect of outlying values in index patients.

The third condition assumes no changes in prognosis have occurred since the time the reference patients were treated. This applies to changes not accounted for by the prognostic factors in the model. If independent changes have occurred, model predictions may be too pessimistic.

At the onset of the two Dutch transplantation MTA's much information was available concerning prognosis in end-stage heart and liver disease. However, despite many efforts to use this information in the case of end-stage heart disease, results were disappointing due to the lack of suitable prediction models (Van Hout, 1988; Bot, 1988A).

In a contrast, analysis of shadow survival in liver transplantation was more succesful. At the time of data collection there was a small number of prognostic models on cirrhosis (Bonsel, 1988F). Also the second condition of prognostic modelling was satisfied. With regard to the last condition, slight differences between older and more recent prognostic models for primary biliary cirrhosis patients suggested a small improvement of conventional treatment results. Nevertheless the use of prognostic modelling seemed justified in the liver transplantation study.

4.3 Technical features of survival analysis

Purposes

Survival analysis may serve three purposes: description of survival times and terminal events, comparison of survival and prediction of survival.

Description of survival may take different forms, depending on the presence of censored observations and the nature of the survival data[11]. Frequently an arithmetic function is derived to approximate parametrically the observed survival-times distribution.

Comparison of survival of different groups is an obvious means to evaluate interventions (be they therapeutic or otherwise)[12]. A number of tests are available. Heterogeneity of the survival-times distribution may also be investigated using regression techniques.

The prediction of (individual) survival requires the application of the above mentioned techniques, particularly Cox' regression. Because of its relevance we shall elaborate on its application.

Techniques for description

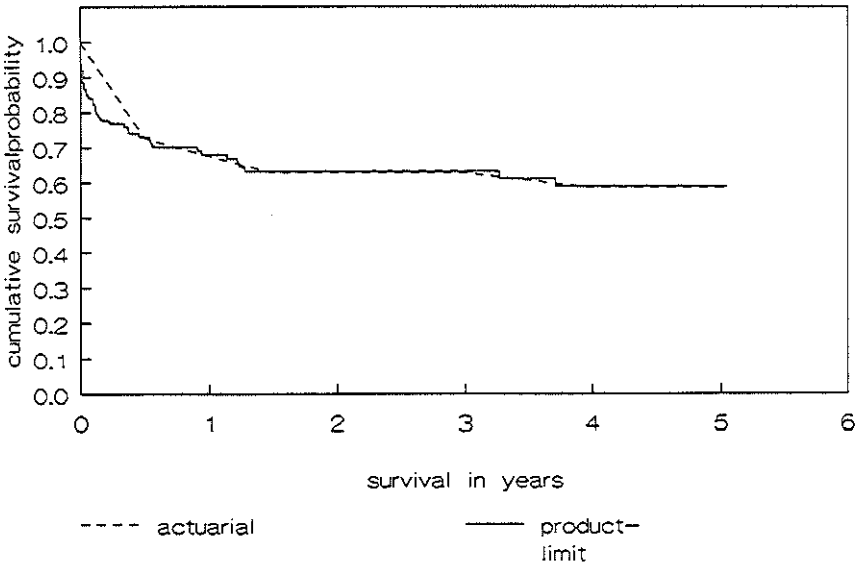
Survival can be described using $S(t)$, $f(t)$ and $\lambda(t)$. These functions represent the cumulative survival function, the probability density function of survival and the hazard rate function respectively. If one of these functions is known, the others can be calculated[13]. If $S(t)$ is known, $LY(t)$ which describes the expected number of years to live after t years of follow-up, can also be derived[14,15]. Survival analysis from the medical point of view uses survival function $S(t)$, from the economic point of view the integral $LY(t)$[16].

If survival data are uncensored, exact cumulative survival probabilities

may easily be calculated for any point of time during follow-up. A special case is the reduced sample technique which calculates cumulative survival probabilities within a specified period, using only data of those included in the study for this period or longer. In clinical studies this is usually not feasible due to the wide range of follow-up periods[17].

Two methods exist which deal with censored data (see Figure 4.1).

Figure 4.1 Survival after liver transplantation - two methods.



If data are aggregated the actuarial or life-table method is used (e.g. in population survival; Cutler, 1958; Benedetti, 1983; Lee, 1980). If individual survival times are known, the product-limit method is used (e.g. in patient survival after liver transplantation; Kaplan-Meier, 1958).

The first method presumes intervals to be defined, usually equidistantly[18]. If each interval contains only one observation, this method is reduced to the product-limit method. Obviously in Figure 4.1 the graphic impression of the actuarial method in the first year is too optimistic, due to the fact that mortality is not evenly distributed along this interval. This illustrates a disadvantage of the actuarial method.

Techniques for comparitive and explanatory analysis

Two or more survival distributions may be compared. Two commonly employed methods are the log-rank test (Peto, 1972), the Cox-Mantel test (Mantel, 1966) and the generalized Wilcoxon tests (Gehan, 1965; Peto, 1972). These tests differ in how observations are weighted. The Wilcoxon tests have more power than the other tests if the hazard ratio is time invariant (Lee, 1980 [section 5.1]; Hop, 1981; Benedetti, 1983).

Comparison may also take another form. Instead of comparing two or more survival-times distributions, the effect of several variables on survival may simultaneously taken into account.

If we are interested in the effect of a set of variables on the survival at a *fixed* point in time, 5 year survival, we may use the logistic regression model and the reduced sample of survival times. This model does not use information on the time of occurrence of terminal events. We did apply the model in analyzing the (potential) effect of donor organ quality on graft survival (Pruim, 1989)[19].

If we are interested in the influence of a set a covariates on *the entire survival-times distribution*, the proportional hazards model may be used (Cox, 1972; Kalbfleisch, 1980; Hopkins, 1983; Lee, 1980).

$$\lambda(t) = \lambda_0(t) * \exp(b_1z_1 + b_2z_2 + \ldots + b_pz_p) \qquad (1)$$

In equation 1 $\lambda(t)$ is the hazard function and $\lambda_0(t)$ is an unspecified base-line hazard function. z_1, z_2, ..., z_p describe the value of each of the p covariates of a subject. The coefficients of the model to be estimated are b_1, b_2, ..., b_p. The coefficients and the base-line hazard function are estimated by means of a maximum likelihood method.

This proportional hazards model may be further refined. First of all the model may be elaborated for stratified analysis when the proportionality assumption is violated[20]. For each stratum i the coefficients are identical, unlike the $\lambda_{0i}(t)$:

$$\lambda(t) = \lambda_{0i}(t) * \exp(b_1 z_1 + b_2 z_2 + \ldots + b_p z_p) \quad (2)$$

An application for transplantation is included in Gilks (1986).

A related extension is the stratification of the covariates. In the two-strata case we get (a is the stratum indicator):

$$\lambda(t) = \lambda_0(t) * \exp(b_{1,a} z_1 + b_{2,a} z_2 + \ldots + b_{m,a} z_m + b_{m+1} z_{m+1} + \ldots + b_p z_p) \quad (3)$$

Variable z_1 takes the value 1 or 0, and a indicates the stratum. Variables z_2 to z_m are variates with a significant different explanatory effect, hence $b_{2,a=1}$ differs from $b_{2,a=0}$. Finally variables z_{m+1} to z_p are variates in the model with an effect independent of the stratum chosen.

In his analysis of cirrhosis patients treated with either azothioprine or prednisone Christensen proved the usefulness of this approach (Christensen, 1986). Theoretically this method is superior to that applied by Neuberger (1989) in estimating survival impact of liver transplantation, accounting simultaneously for prognostic factors which influence non-transplantation survival, transplantation survival, or both. However, its data requirements are considerable[21].

A last extension of the model may be the incorporation of time-dependent variables. Equation 4 describes its general form.

$$\lambda(t) = \lambda_0(t) * \exp(b_1z_1 + ... + b_mz_m + b_{m+1}z_{m+1}(t) + ... + b_pz_p(t)) \quad (4)$$

The function z(t) may be any function but usually it takes the following 'switch' form (a=1, b=0, or the reverse) (Breslow, 1975):

$$z(t) = \begin{cases} z(t) = a; t \leq z \\ z(t) = b; t > z \end{cases} \quad (5)$$

A refined version of this time-dependent indicator variable approach was suggested by Cox (1984 [p.129]).
A relevant example of this time-dependent application of the proportional hazards model using a 'switch'-variable may be found in Van Hout (1990). In cirrhosis no examples seem to exist[22].

So far we are not aware of any application of the following more general elaboration of *z(t)*:

$$z(t) = f(t;z) \quad (6)$$

Generally, this extension of the proportional hazards model seems promising in many pathological processes. It enables to take the *change* over time of prognostic variables into account, rather than using the level of a prognostic variable at the onset of the observed survival time (as in the common time-fixed proportional hazards model). For example the application in cirrhosis seems particularly justified.

Techniques for prediction

The covariates of the proportional hazards model may be interpreted as *prognostic* variables. If the model is estimated and $\hat{b}_1$, $\hat{b}_2$ ... $\hat{b}_p$ and $\hat{\lambda}_0(t)$ are known, for each individual j we may calculate $\hat{b}_i z_{ij}$ as a prognostic index ($\hat{PI}_j$):

$$\hat{PI}_j = \ln(\hat{\lambda}(t)/\hat{\lambda}_0(t)) = \hat{b}_1 z_{1j} + \hat{b}_2 z_{2j} + \ldots + \hat{b}_p z_{pj} \qquad (7)$$

Several approaches exist to estimate the variance of $\hat{PI}_j$ (O'Quigley, 1983; Andersen, 1983; Christensen, 1987). Van Houwelingen (1990) introduced alternative estimates of the error of $\hat{PI}_j$. If $\hat{PI}_j$ and the estimated cumulative base-time hazard function $\hat{\Lambda}_0(t)$ are known, the function $S(t;PI_j)$ of patient j predicts individual survival probabilities as follows:

$$\hat{S}(t,\hat{PI}_j) = \exp(-\exp(\hat{PI}_j \hat{\Lambda}_0(t))) = \hat{S}_0(t)^{\exp(\hat{PI}_j)} \qquad (8)$$

We define a prognostic model as a function which includes a basal hazard function and a set of weighted prognostic variables and which together predict survival probabilities in an individual case.
Note that $\hat{\Lambda}_{0k}(t)$ is different for each prognostic model. (The index k refers to the particular prediction model chosen.) Hence when $\hat{PI}_{Aj}$ and $\hat{PI}_{Bj}$, based on model A and model B respectively, have an identical value, this does not imply identical estimated survival funtions[23].

The aggregated survival function $\hat{S}(t)^*$ of a group of n subjects may be computed as follows. Suppose a group of n subjects, each subject j with an individually predicted survival function $S(t,PI_{kj})$ estimated by:

$$\hat{S}_0(t)^{\exp(\hat{PI}_{kj})}$$

For each subject j, either dead or alive at the final follow-up date, the maximum possible follow-up time $t_{max,j}$ is defined as the difference between final follow-up date and date of entry in the study[24].
Maximum possible follow-up times are put in reverse order of date of entry, the most recent first, resulting in an ascending order of $t_{max,j}$:
$t_{max,1} < t_{max,2} < \ldots < t_{max,n}$.
We define n corresponding time intervals I_m (m = 1, 2, ..., n), with in each interval one patient less in possible follow-up, as follows:
$I_1 = [0, t_{max,1}>$, $I_2 = [t_{max,1}, t_{max,2}>$, ... , $I_n = [t_{max,n-1}, t_{max,n}>$.
Next, the survival function for each individual j is computed from t=0 to $t_{max,j}$. For each time interval I_m, the aggregate nontransplantation survival is computed as the average survival of m patients with a follow-up including I_m, as follows:

$$\frac{1}{m}\sum_{j=1}^{m} \hat{S}(t,\hat{P}I_{kj}) \qquad (9)$$

The index m represents the number of patients with $t_{max,j}$ exceeding the upper limit of I_m, which decreases from n patients in I_1 to one in I_n.
Next conditional survival probabilities are derived for each period I_m, each time assuming complete survival at the start of that period.
The estimate for the group survival function $\hat{S}_k(t)^*$ for the entire interval I $[0, t_{max\ j=n}>$ is subsequently obtained by multiplication of all conditional period survival probabilities, thus linking all adjacent intervals I_m. This procedure yields a survival curve which is comparable to the product-limit estimator of observed survival, because expected individual survival functions are censored at the date of $t_{max,j}$.

Two alternative approaches were proposed by Neuberger (1986) and Markus (1989) respectively. The Neuberger-formula of the aggregated survival curve for the entire interval I reads (see next page):

$$\hat{S}_N(t)^* = \hat{S}_{0N}(t)^{\exp(\frac{1}{n}\sum_{j=1}^{n} \hat{PI}_{Nj})} \quad (10)$$

The Markus-formula of the aggregated survival curve for the entire interval I reads:

$$\hat{S}_M(t)^* = \frac{1}{n}\sum_{j=1}^{n} \hat{S}_M(t,\hat{PI}_{Mj}) \quad (11)$$

The Neuberger-formula is particularly erroneous as it denies the exponential nature of $\hat{PI}_j$ with regard to $\hat{S}(t,\hat{PI}_j)$. The Markus-formula resembles ours, except for a differently defined risk set. We illustrated the magnitude of these biases in a PBC-population (*A3*).

4.4 Survival analysis in liver transplantation - overview

During the development of liver transplantation an RCCT has never set up, despite the high mortality of the procedure in the first decade of its clinical application. Remarkably, the explicit analysis of the survival *benefit* of liver transplantation as opposed to best conventional therapy it not to be found in the literature before 1988, 25 years after its introduction.

In the pioneering stage little survival information was made available. In the early adoption stage of liver transplantation survival analysis consisted only of crude mortality figures. Diagnostic categories usually were combined, selection of reported cases often was often unclear and standard survival techniques were employed infrequently (Bonsel, 1988F).

Since the NIH-conference in 1983, which resulted in diffusion of liver transplantation, the quality of reports has gradually improved (Anonymous, 1983A & 1984).

Interest in factors predicting survival after liver transplantation has also been growing. This is clearly related to the availability of statistical packages containing routines for proportional hazards model estimation. Despite the awareness of the simultaneous effect of predictive factors on the survival with and without liver transplantation - the 'timing' problem (Haagsma, 1990 [p.5]) - no formal analysis on the survival *gain* with liver transplantation was published before 1988.

In 1988 the final report of the Dutch liver MTA appeared, containing formal analysis of the net benefit of transplantation with regard to survival and quality of life. In this context we should also mention two articles on survival gain of liver transplantation in PBC-patients in 1986 (Neuberger, 1986) and 1989 (Markus, 1989) respectively. As these research groups in Birmingham (United Kingdom), Rochester (United States) and Rotterdam were mutually unaware of their developmental work, the coincidence is remarkable.

4.5 Survival analysis of liver transplantation: a selection of results

In this section we shall present some results of the Dutch liver transplantation MTA. Description of the programme and the patients as well as details about the survival analysis are provided elsewhere (Bonsel, 1988E,F, 1989B; *A3, A10*).

The use of *formal expert judgment* was discussed with the transplantation team of the AZG in Groningen. However, this method was regarded to be unfeasible for various reasons.

The straightforward use of *historical data* was considered in liver carcinoma, but due to the small number of transplanted cases (3), a similar analysis already published was regarded to be sufficient (Ringe, 1989).

The *quasi-experimental* approach was used in the analysis of biliary atresia. Children with biliary atresia were treated with liver transplantation in the Dutch centre in Groningen from 1982 onwards. In the following 5 years only a fraction of these patients were referred to the AZG for liver transplantation in a aselect way (probably less than 25%). Most patients were treated elsewhere with a Kasai procedure only.
Houwen published the combined Dutch experience in 1972-1987 with this treatment strategy (Houwen, 1989)[25]. Figure 4.2 compares the Houwen data with results up to 31-12-1989 in the AZG.

Figure 4.2 Survival in biliairy atresia - Kasai alone (1977-1987) vs. Kasai and LTx (1982-1987). The Dutch experience.

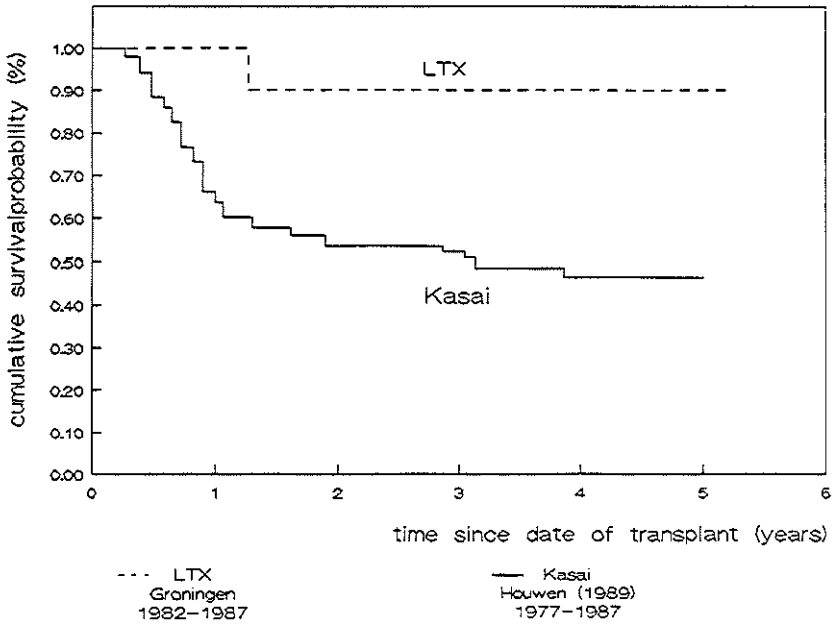
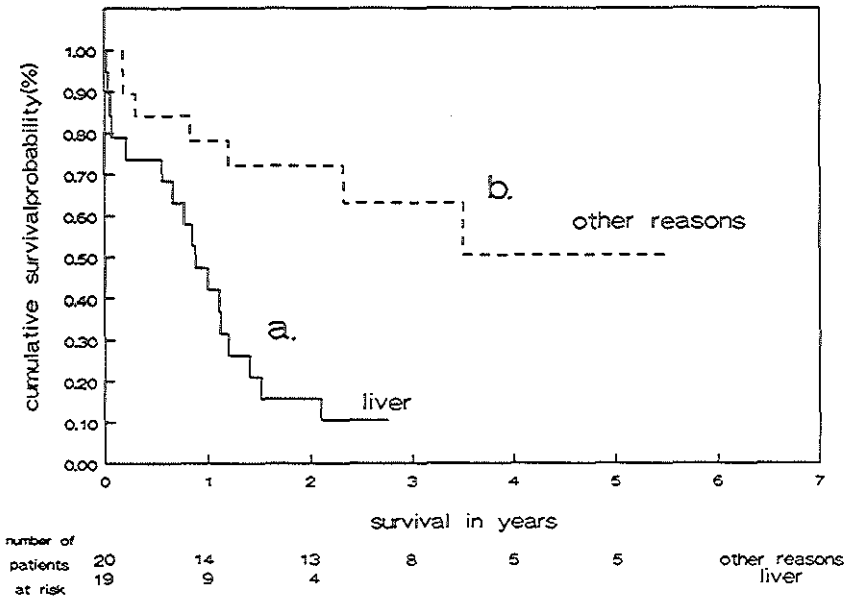


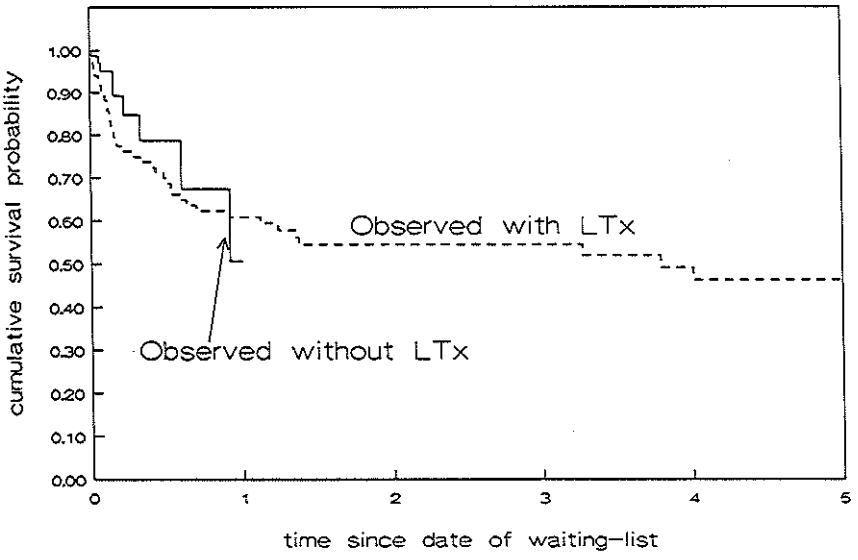
Figure 4.3 Survival after screening for liver transplantation. Patients rejected for liver dysfunction (a.) and other reasons (b.)



The analysis of rejected patients was also carried out. Patients with hepatobiliary malignancies were excluded as the course of their disease is different from that observed in chronic cirrhosis. Figure 4.3 presents the survival of patients rejected for severe liver dysfunction precluding transplantation (a., n=19) and for other reasons (b., n=20)[26].

Though a. by definition represents a pessimistic average shadow survival of a transplanted cirrhotic patient, it provides us a useful lower boundary of non-transplantation survival. In retrospect b. was rather close to the predicted non-transplantation survival curves of the group of patients with PBC or non-PBC cirrhosis.

Figure 4.4 Survival from entry on waiting-list. Transplantation period excluded ("without LTx") vs. transplantation period included ("with LTx")



Waiting-list analysis has been rather unsuccesful in the Dutch liver transplantation MTA. We investigated waiting-list survival, with three definitions of t_0. If t_0 was defined as date of entry in the screening protocol or as date of provisional candidacy, the assumption of homogeneity at entry clearly could not be met.

If t_0 was defined as date of appearance on the Eurotransplant waiting list for livers, homogeneity seemed reasonably guaranteed, but due to rather short waiting times, a valid shadow survival curve could be derived for the first year after liver transplantation only (see Figure 4.4). For this reason waiting-list analysis in our case was only of limited relevance[27].

Additionally we applied the proportional hazards model, starting at the three t_0's mentioned before, and including one time-dependent variable (i.e. transplantation date). None of the models showed prognostic significance of transplantation (Bonsel, 1988F). In the case of the model with t_0 defined as date at entry on the Eurotransplant waiting list, this result could have been predicted from Figure 4.4.

Finally we applied *prognostic modelling* in cirrhotic patients. First we shall give some general information on cirrhosis models. Then one application in non-PBC cirrhotics will be presented.

Before the availability Cox' regression model, the prediction of the course of cirrhotic disease (particularly of alcoholic origin) received already considerable interest. Ratnoff already pointed to the unfavourable signs and symptoms of cirrhosis whose influence now has been estimated more precisely (see below; Ratnoff, 1942).

Until the late 1970's, survival studies suffered from the unavailability of suitable statistical tools (see e.g. Garceau, 1963; Hällèn, 1963; Powell, 1968; Child, 1968 [p.49]). Siegel's multiplicative risk model - in present statistical language - was a remarkable achievement (Siegel, 1969), though we are unaware of application of its cumbersome calculus outside the primary developing centre.

At present there are many prognostic models for cirrhosis (Porayko, 1990; Christensen, 1985 & 1989; Llach, 1988; Ginès, 1987; Ji-yao, 1989; Dickson, 1989; Wiesner, 1989; *A3;* see also Table 4.3). Cross-validation experience has so far been limited (*A3*; Grambsch, 1989).

A striking similarity may be observed. The cirrhosis models contain at least one variable related to the hepato-renal syndrome (ascites), one related to the synthesis capacity of the liver (albumin, PTT), one related to cholestasis (bilirubin) and one frailty variable (age).

As might be expected, the cholangitis model is slightly different.

Table 4.3 Seven prognostic models for cirrhosis

Author	*Diagnosis*	*Prognostic variables*
Christensen (1985)	PBC	AG,AL,AZ,BI,PA
Dickson (1989)	PBC	AG,AL,AS*,BI,PT
Christensen (1986)	CAC,CIC,ALC	AC,AG,AL,AP,AS,BI,GB,NK,PA,PT
Bonsel (*A3*)	PBC,CAC,CIC	AG,AL,AS,BI,EN,GB,HS,IC,PT
Ginès (1987)	CAC,CIC,ALC	AG,AP,BI,GG,HS,PT,SX
Ji-yao (1989)	PBC,CAC,CIC, ALC	AG,AL,AS,BI,EN,GB,HB,PT
Wiesner (1989)	PSC	AG,BI,HB,IB,PA

AC = alcohol consumption
AG = age
AL = albumin
AP = alkaline phosphatase
AS = ascites (NB AS* = edema)
AZ = azathioprine treatment
BI = bilirubin
EN = encephelopathy
GB = g.-i. bleeding or varices
GG = gamma-globulin
HB = hemoglobin
HV = HBsAg +
IB = inflammatory bowel disease
IC = icterus
NU = nutritional status
PA = pathology liver
PT = prothrombin time
SX = sex

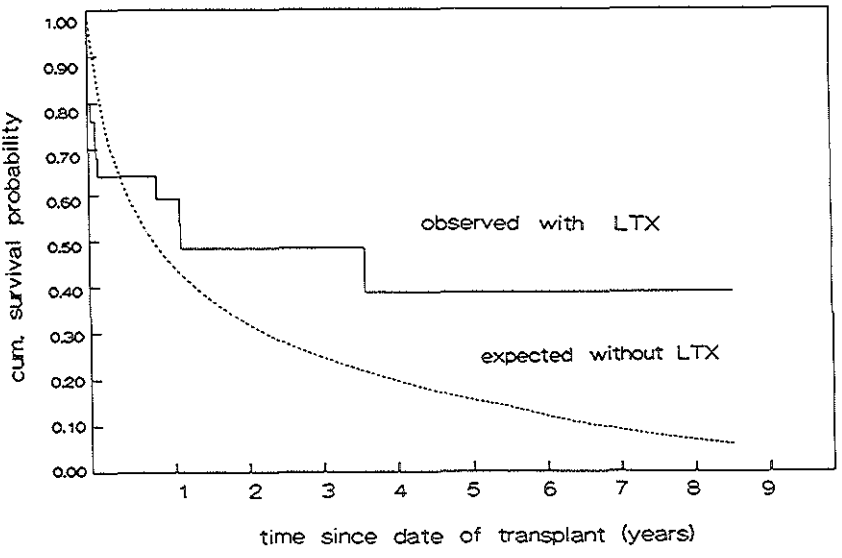
These empirical models may be compared with the still universally used prognostic Child-Pugh classification (Child, 1968; Conn, 1981).

The *choice of the prognostic factors* of the Child-Pugh classification may be criticized. The omission of age does not seem to be justified. The inclusion of hepatic coma is probably redundant; death usually rapidly follows hepatic coma as Hippocrates already observed in his Epidemics (1983B [p.136-137]). Hence hepatic coma has little prognostic relevance in our case. The *efficiency in prediction* is inferior to that in the prognostic models for two reasons. The crude practical interpretation (three categories instead of weighted real values) implies loss of information.

The assumption of equivalence of factors (unweighted summing up instead of weighted summing up, or better, multiplication as in Cox' model) is empirically untenable (Christensen, 1984; Ji-yao, 1989).
However, as may be observed with similar rules in other diseases, the apparent sophistication of Cox' models (see e.g. Haagsma, 1990 [p.5]) has not challenged the position of the Child-Pugh classification.

Next the results are shown which resulted from application of Christensen's non-PBC model to 25 transplanted AZG-patients (Figure 4.5; Bonsel, 1988F). The results in PBC are published elsewhere (*A3*).

Figure 4.5 Observed and model-predicted survival in non-PBC cirrhotic patients. Prognostic model of Christensen



Particularly in the case of the non-PBC model of Christensen the validity of the application was investigated.

The Christensen model combines experience from alcoholic and non-alcoholic patients. It assumes no difference in prognostic factors apart from the intake of alcohol. At our request, Christensen kindly reviewed model estimates for alcoholic and non-alcoholic patients separately: the relative importance of prognostic factors appeared to be essentially the same, apart from the additional alcohol factor in alcoholic patients (E. Christensen, personal communication).

Histological re-examination of all Christensen's non-alcoholic patients was not feasible, hence we do not know the exact distribution of etiological entities (particularly the auto-immune fraction). Our own analysis of a (limited) number of Dutch patients (131) with non-alcoholic cirrhosis of all kinds did not reveal a separate influence of diagnosis, hence we think that potentially a different mix of non-alcoholic cirrhotic patients does not violate the diagnostic homogeneity assumption.

The homogeneity of prospect assumption seems slightly violated. The reference population of Christensen's model dates back from the early sixties. In the last 16 years Ginès and others have demonstrated the beneficial effect of rigorous treatment of the hepato-renal syndrome (Ginès, 1988). Their evidence suggests that a better overall survival in decompensated patients may be currently achieved. We expect the prediction presented in Figure 4.4 to be the lower boundary of shadow-survival in non-PBC cirrhotic patients who did reveive a transplant.

Finally a technical problem in the application of the Christensen-model should be mentioned. Laboratory methods have changed, so not all these type of variables in the model are measured by the same technique. Though in some variables comparibility appeared to be sufficient, it remained uncertain in others (e.g. serum alkalic phosphatase).

4.6 Summary

Survival analysis in MTA essentially consists of the application of already available statistical tools. Supplementary calculations are necessary to arrive at life years gained as an economic unit of effectiveness.
Full exploit of explanatory regression techniques may be helpful in the estimation of survival consequences if no RCCT information is available. From our experience in liver transplantation we conclude that this approach is possible provided one or more of the following conditions are met: a. quasi-experimental control group data are available, b. intervention delay data are available, c. historical control group data are available, d. prognostic modelling is feasible using data from a., b. or c..
In all cases homogeneity of reference and index patients with regard to diagnosis and prospect should be sufficient, although the last method provides opportunities for adjustment. Particularly for prognostic modelling extensive data collection is required of both index and refernce patients (see e.g. Bonsel, 1988D).

5 ANALYSIS OF HEALTH STATUS

5.1 Introduction

This chapter presents some considerations on the analysis of health status.

A separate discussion of health status* is justified for several reasons. At the theoretical level the debate about how to measure health status still continues. At the practical level different choices on the design of health status measurement hamper the comparison of results, which is a major disadvantage in the context of MTA (see chapters 2 and 3). The commercial aspects of health status measurement add a powerful dimension to these scientific discussions. Lastly the growing public concern about the meaning and the application of health status measurements can no longer be ignored[1].

The chapter is divided into two parts, parallel to the two concepts of patient outcome distinguished in section 3.4: a medical and an economic one.

The medical concept of patient's outcome consists of three independent aspects (see section 3.4): survival, non-disease-specific health status, and disease-specific health status, of which the first two usually are the most important[2]. Survival was discussed in chapter 4; in the first part of this chapter attention will be paid to non-disease-specific health status.

As explained in chapter 3, the economic concept requires a single measure as an expression for the overall value of the outcome (or 'utility'); generally money as an indicator for value is not accepted, hence

---

* 'Quality of life' and 'health status' have different connotations, 'quality of life' being the more general term. In the context of measurement of health-related aspects of life we have a slight preference for 'health status', the term which will be commonly applied in this text.

proxy measures have been proposed such as life-years and QALY's. The use of life-years alone as a simple extension of conventional survival analysis was illustrated in chapter 4.
In the second part of this chapter composite measures like QALY's are discussed. These composite measures combine information on survival and health status.

Proceeding along these two concepts, we present a brief account of the concept, of the conceptual position of 'time', of the methods of operationalization, of the measurement instruments currently in use and of the methods of data collection.

Reference will be made to the investigations undertaken during the transplantation MTA's which are largely presented in the appendix (*A4 - A8*). We have tried to avoid duplication of information.

5.2 Description of health status - the medical concept

Descriptive health status: a theoretical, unobservable concept

Health status is a theoretical concept. It is not observable straightforwardly like weight, colour, temperature, velocity or, in the domain of medicine, blood pressure and heart rate. As a concept this state like 'stress', 'frustration', 'self-esteem', etc., defines health status to be an abstract mental construct, without a simple unambiguous empirical representation. The various words in use to describe its indirect measurement point to this ('indicator', 'index', 'scale').
As a result the measurement of health status and the subsequent statements on its value rests upon convention. This convention includes a theoretical structure (nomological network) in which health status

occupies a specific position. This structure in turn defines relations between the included concepts (among other things health status), it may prescribe instruments to measure concepts which are defined as measurable and lastly it enables predictions about expected relations between the collected data. From this follows that the theoretical structure also represents *what* the investigator thinks health status is, and subsequently how *validity* of the data may be established.

Clinicians usually do not feel at ease if they are confronted with this abstract account of the concept of health status. However, they are usually unaware of the multitude of such unobservable concepts in clinical practice. Examples are heart function, liver functional capacity, neonatal condition[3]. These concepts have much in common with health status from a methodological point of view.

As with health status, these 'natural' clinical concepts give rise to a number of approximating measurement instruments, which have been developed parallel to the development of the concept.

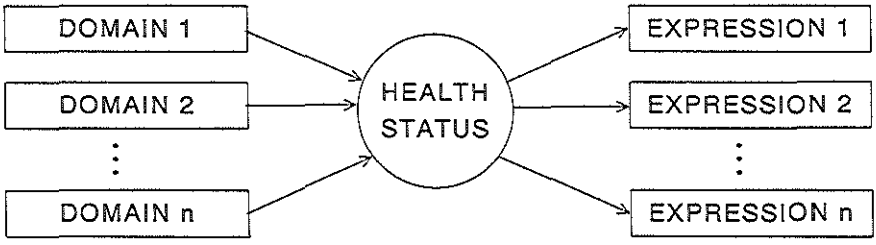
A second parallel is the continuing debate about the best operationalization of these concepts e.g. about the best indicator of liver function. As with health status, the debate is reigned by both a priori scientific beliefs and the results of empirical investigations.

Descriptive health status: a theoretical structure

We regard health status to be a characteristic of human life, defined by aspects such as grief and misery and by satisfaction and happiness. Since there is usually no straightforward relation between the theoretical concept 'health status' and particular empirical observations, a theoretical structure should be devised. A general form of this structure is presented in Figure 5.1 which is derived from Campbell (1976 [p.16)]).

In the centre we find health status, the core concept which lacks direct empirical representation. On the left we find the so-called *domain* part (instead of domains some speak of dimensions or aspects), on the right there are what we might call *expressions*. (Refinements of Campbell with regard to the domain part are omitted).

Figure 5.1 A theoretical structure for descriptive health status



Domains (aspects, dimensions) are thought to determine health status. Domains include physical, psychological, social and various other domains. Considerations about their selection are found in De Haes (1985, 1988) and *A4*. Domains refer to attributes of an individual which may be

determined objectively like walking distance and depressive status (!)[4].
It is a matter of taste and convention whether or not to include non-health related (or not unequivocally-health related) domains. Expressions are more difficult to define. Usually concepts are used which are thought to be related causally to health status. Experienced well-being, both in an affective (happiness) or cognitive (satisfaction) way are the most common concepts encountered. These two concepts have their background in psychological theory[5]. We might also think of a specific behavioural response or more remote consequences of health status changes.
A clinical example may illustrate the structure. If we read liver function instead of health status, 'expression' might be read as (e.g.) hepatorenal syndrome and 'domains' might be read as ascites-score, serum-albumin and renal excretion of sodium under standard conditions in that case.
Such a structure may be used to analyze the health status concept in use in a particular study and also to analyze a particular measurement instrument.
At this point we must distinguish between two situations encountered in health status analysis.

In the first situation the theoretical structure provides us with expressions which may be defined as causal consequences of health status changes or as the representation of these changes. In the structure proposed by Campbell, the latter type of expressions are formed by the particular responses of (healthy) individuals to questionnaires representing 'happiness' and 'overall life satisfaction'. The responses to these questionnaires are thought to represent a direct consequence of the unobservable, latent variable 'quality of life'. Other measurable consequences may be thought of.

In the second situation expressions are lacking: the concept of health status lacks consequences to be predicted (which does not preclude the existence of predictive properties).

The hepatorenal syndrome example may be regarded as a clinical representation of the first situation; the NYHA-classification and the APGAR-score probably are clinical examples of the second. In both situations the choice of the domains origin from a priori reasoning (see *A4*), but in the first case opportunities for validation are better as pointed out by Kirschner (1985). For further conceptual discussions we refer to the extensive literature available[6].

Descriptive health status: the position of time

In spite of the fact that 'time' is not a variable or an element in the model dsecribed above, it has a definite conceptual position in descriptive health status measurement. The concept of health status, in the sense in which it is used in MTA and other evaluative research, is defined as a permanent attribute of human life. Thus, descriptive health status measurement aims at efficiently describing a concept which, like body temperature and heart rate, is defined permanently but which, unlike these examples, lacks continuous monitoring devices.
Approximative measurement of health status is therefore necessary, which usually implies that patients are asked to participate in interviews at regular intervals. The frequency of interviewing and the period to which health status questions refer are important conceptual choices.
Overall two schedules may be distinguished, a before-after measurement design and a longitudinal design[7].

If a unique, lasting change results from intervention, two measurements appear to be sufficient, one before and one after intervention.

The post-intervention measurement should be scheduled at the time the recovery process may be expected to be finished. This design is tempting for two reasons: it is usually easy to implement and standard statistics suffice to calculate differences (and e.g. to estimate a priori an appropriate sample size).

This classical experimental schedule is often not justified.

First, the pace of changes in health status may vary. For example in the heart transplantation study the immediate changes in cardiac function preceded changes in mobility and energy (return to almost normal levels after weeks), which in turn preceded changes in role function and work situation (return to normal levels after months to years; Bonsel, 1990B).

Secondly, information on the course of disease and not merely the status at a particular endpoint is of critical value. Measurements during the treatment episodes give valuable information about the price to be paid for ultimate recovery. Longitudinal information will also shedd light on the process of psychological adaptation.

Thirdly, definite and complete recovery without recurrence or change is a desirable but rarely encountered endpoint of a disease process.

For these reasons longitudinal designs should usually be preferred, although they carry with them specific problems with regard to feasibility and type of analysis (particularly the handling of losses to follow up).

In practice the frequency is not the result from conceptual considerations alone: clinical practice and the condition of the patient usually pose some limitations.

Generic health status measurement instruments: conceptual considerations

Given the health status model, its operationalization (i.e. the construction of measurement instruments and the data-collection design) should

be decided on. Dessens and Janssen (1987) already pointed to the relevance of different operationalization methods for the health status model[8]. Though this may be relevant for the further development of health status research, the issue will not be discussed here. Attention will be restricted to disease-non-specific (syn. generic) descriptive health status instruments which result from the standard approach[9].

Three types of generic health status measurement instruments may be distinguished according to the ascending level of aggregation: a classification, a profile and an index[10]. A special feature of a profile is that if it claims to be comprehensive an index may be developed validly from a profile. Thus we constructed an NHP-index, based on the responses of about 100 renal dialysis patients (Bot, 1991).

For a discussion on other features of generic health status measurement instruments like the particular dimensions, items and response modalities to be chosen, the reader is referred to current textbooks and recent publications on health status measurement in MTA (*A4*, Bot, 1991). The criteria feasibility, internal consistency, reliability, validity, responsiveness/sensitivity, suitability to support patient outcome valuation (in the context of the economic concept of outcome - section 3.4) and suitability for use in the Netherlands enable judgment on current instruments. Surprisingly, basic psychometric evidence on the properties of well-known instruments is rather scarce.

Measuring health status: practical considerations

Basically three techniques are available for data collection: observation, questionnaires and interview. Each method has its advantages and disadvantages. Indirect measurement by observation should be reserved for special occasions (e.g. the very young or the mentally handicapped).

The main advantage of questionnaires in this context are their modest use of resources and their ability of eliciting response in a standard way. A major advantage of the interviewing method is the possibility of the interviewer to enhance data quality during the interview and to record additional data (characteristics of the patient, for example the response behaviour). However, patient interviewing is demanding. To obtain reliable standard health status scores requires the availability of experienced interviewers, a sophisticated interview design, an equally well-developed design of the answer scoring and ample resources. Given these characteristics we usually prefer patient-based questionnaire data-collection in MTA.

Regardless of the data-collection method, a closed and structured design is required to achieve a standard and quantitative health status-description. However, sometimes additional research objectives may suggest the supplementary use of an open, less structured design. E.g. in the MTA of the British Heart Transplantation Programme the administration of one highly structured questionnaire (NHP) was combined with a semi-open interview to ascertain that no unexpected consequences of this new technology remained unnoticed using the standard questionnaire alone (Buxton, 1985).

Lastly we should mention computer assisted interviewing (CAI), a technique of data collection which has become available recently. The computer may be used by the interviewer or the respondent.
In the latter case the computer presents a questionnaire to the patient on a monitor. The patient uses the keyboard to express his opinion: alphanumeric keys to 'write' open answers, numeric keys to choose from predefined answers and special keys for special features e.g. the drawing of lines (visual analogue scale). Up to now CAI has proven to be a valid substitute for paper and pencil questionnaires. Our own experience with

various patient groups has shown CAI to be advantageous in a research setting with regular follow-up at a clinical or an outward department (Bonsel, 1988).

Description of health status in MTA's of solid organ transplantation

Below we give a short account of the strucure of health status analysis in the MTA's on which this thesis was based, and of the two predecessors of the heart tranplantation MTA, by Evans (1984) and Buxton (1985) respectively.

Campbell's model was completely adopted by Evans in the MTA of heart transplantation and the MTA of treatment modalities for patients with end-stage kidney disease (both studies took place in the United States). In these studies the separate Sickness Impact Profile dimensions represented the domains. The so-called Index of Psychological Affect scores and the Global Well-Being scores represented the expression part of the model[11]. No evidence was published about the fit of the observations to the proposed structure.

Evans adopted a cross-sectional measurement design probably for practical reasons. This design limited the interpretability of the results (for example a 'healthy survivor' effect was likely; Bonsel, 1988).

Buxton applied the general structure without particular expression concepts. A patient interview with specific attention to a priori defined domains was combined with a fixed 3-month administration of a health status questionnaire (the Nottingham Health Profile).

In the Dutch MTA's of heart and liver transplantation Campbell's model was adopted, adding some of his refinements of domain measurement. As in this particular medical context we regarded the future outlook, as perceived by the patient, an essential part of the expression

side of the model, a measurement instrument for this concept was developed (Bonsel, 1988B & 1988C).
The fixed 3-month time schedule was identical to that of Buxton, though some parts of the lenghty questionnaire were for obvious reasons repeated at a lower rate (Bonsel, 1988B). Computer asisted interviewing was applied. Results are published elsewhere (*A4*, Bonsel, 1990B). These studies too have not reported on the confirmation of Campbell's model.

5.3 Valuation of health status - the economic concept

With the economic concept of health status we enter the domain of ends and means in health care. The means essentially are the health services i.e. the technology under investigation or its best alternative. The ends refer to patient outcome.
As explained in chapter 3, the price of health care services does not reflect the economic value of the associated patient outcome. Health economists have accepted other parameters which all are thought to be relevant outcome parameters: life years gained, quality adjusted life years gained and the healthy years equivalents gained[12]. All are based on length of life as the core of patient outcome, which seems reasonable from a conceptual point of view (see Richardson, 1990 [p.16-18]).
The straightforward use of life years alone (in conventional cost-effectiveness analysis) requires an extension of standard survival analysis (see chapter 4). Below we shall concentrate on the more complex procedures for valuation of patient outcome, which are needed in cost-utility analysis. The section starts with a general account of utility. As Richardson (1990) has published a highly estimated analytical paper on the subject, reference to his paper will be made repeatedly.

The economic value of patient outcome: a theoretical concept

In the preceding text we have loosely used the term 'economic value' or 'utility'. An indication or, better, a definition of utility is necessary to arrive at a suitable operationalization. This definition is to some degree a matter of convention, as utility occupies a specific, but not exactly defined, position within economic theory. Richardson's account of this position shows that the following two definitions (interpretations) are the most important in our context:

1. utility is an index which represents the strength of choice or preference, with cardinal properties;
2. utility is the linear summation of probabilities times quantities that are defined by a set of axioms.

The first definition describes the concept and states an important property (cardinality) of the index which should be constructed; the operationalization is left open.

With the last definition utility is only operationally defined, though its meaning may be understood within the context of the Von Neumann-Morgenstern axioms of rational behaviour: it is regarded to reflect preference given axiomatic rationality[13].

An important difference between the definitions is the position of 'risk' and/or 'uncertainty', which is an essential element of the second definition.

Next we proceed with utility, however defined, in the context of valuation of patient outcome.

First it should be stressed that *patient outcome* in this section refers to length of life and to all the attributes of health status that are considered to be of additional relevance in decisions (see section 3.4). What we intend to value is an array of outcomes with different lengths of life and

varying levels of health status within each 'branch' (see the vector concept of Mehrez, 1989), which altogether represents the expected future of a patient given some decision.

Secondly, the concept of *risk* underlying the second definition of utility is not equivalent to the concept of uncertainty, which is at issue if we state that uncertainty exists about patient outcome as reflected by the array of possible outcomes (for a detailed analysis, see Richardson, 1990).

Thirdly, in this context the concept of *utility* reflects preference as defined before. However, the preference with regard to particular patient outcomes is unobservable in practice[14]. As stated before, no choice situations exist in which behaviour may be observed which may validly be interpreted as revealed preference: we for example do not need to buy particular medical services priced according to their assumed effect on patient outcome. The comparison of patient outcomes by patients themselves might at least offer a second-best ranking. However, the few instances that a patient faces a realistic choice situation (for examples see the literature on clinical decision making) do not yield the preferences we want to know in MTA: the alternatives are for example limited to those accidentally imposed by the clinical problem. Thus we conclude that, though the theoretical concept of preference seems rather simple, it is still unobservable in a straightforward way.

Patient outcome valuation: the position of time

Time has a double meaning in patient outcome valuation:

1. length of survival is an aspect of outcome associated with a probability to occur and level(s) of health status.
2. the ultimate value of an outcome is expressed in time units (QALY, HYE).

This double meaning is one of the causes of the intricacies of valuation studies: time often appears in one way or another in the health scenario to be valued and in the valuation procedure itself.

Measurement procedure and measurement instruments

A measurement procedure has been accepted for a third-best (see previous alineas) three-stage measurement of preferences (*A5*).
First, a *health scenario description* is derived from prior information on the prognosis (defined as the array of possible survival and health status consequences) of the particular intervention. For pragmatic reasons a health scenario will usually be reduced to the description of a health state with a particular duration and with a fixed particular health status[15]. This predefined duration may reflect complete survival within a particular branch or a period within this branch with assumed constant health status e.g. a year[16]. This index health scenario will be valued in the following step. One of a dozen available methods is used to arrive at this *scenario-specific value* (amongst others category rating, magnitude estimation, time trade-off, standard gamble, see *A4, A5*)[17]. All methods require the allocation of a value to an index health scenario with particular length of survival and with a fixed particular health status, by comparing it with another reference health scenario or with two other reference scenarios. The method of comparison is the assignment of a rank, a value-equivalent, or a probability-equivalent respectively (for details see Torrance, 1986; Bonsel, 1988C; *A6*). The value which results from the comparisons usually is transformed to a value ranging from 0 (zero) to 1 (one). Using modelling techniques a multi-attribute utility function may be derived[18] to predict values of any composition (*A6*).
In the third stage *the transformed value is used as an adjustment factor to*

adjust the life-year value of the predefined period. If the target health scenario already has the complete duration of the branch, this will suffice. If not, additional valuations of the remaining periods within the branch are necessary.

Finally *summation of the resulting adjusted period values* results in a branch-value expressed in quality-adjusted life years (QALY's). Multiple branches may be summed[19] to get an overall value of one path of the decision problem (for example a treatment alternative).

The central assumption of this three-stage procedure is that preferences for different outcomes are approximated adequately by the use of quality-adjusted life year values. As opportunities to test this assumption are scarce, some remarks which arise from the measurement procedure itself are presented.

1. To our knowledge until now never a range of branches but only a single branch with fixed duration (without further survival-ramifications) and a fixed level of health status has been valued. It has been stated that this reduction disposes the real preference problem of the uncertainty aspect (Richardson, 1990). The so-called healthy-years-equivalent method (see Mehrez, 1989A & 1989B) has been devised to adress this problem. This procedure arrives at QALY's (which are called HYE's) using a standard-gamble based method which has been claimed to be able to value a health scenario consisting of *multiple* branches simultaneously. The reported evidence is limited to the singular branch case. As there is no clear conceptual difference between the two-stage HYE-method and the existing one-stage TTO-method (see Richardson, 1990), we prefer the latter or any of the other valuation methods available for their practical convenience.

   No experience has been reported with the valuation of complex patient outcome descriptions; we also used the single branch approach.

2. An important feature of the method of translating the complexity of patient outcome into assessable scenarios is the applied system of health status description. Within a the EuroQol-group, we were able to devise an internationally applicable descriptor system, which partially drew on the earlier experience in the Dutch MTA's (Bonsel, 1988C; *A6 - A8*).
3. A last feature of the first stage of the measurement method is the choice of an unambiguous method of *describing a scenario with a particular descriptor system (that of the EuroQol-instrument**) using empirical data about the health status*. In the Dutch MTA's we adopted a formal approach for the descriptor system developed prior to EuroQol (as described in *A4* and *A6*) but a similar algorithm for EuroQol has yet to be established.

For remarks on the second stage of the procedure we refer to Lipscomb (1982), Froberger (1989), Richardson (1990), and the appendix.
We agree with Richardson that within the context of the second-best methods the time-trade-off method conceptually fits best to the use of units time to represent preferences.
The most difficult challenge of the second stage of this procedure was getting respondents to perform a valuation task that indeed represents a cardinal preference of realistic scenarios. The EuroQol-group showed progress in this respect after lengthy experimentation with the measurement procedure (*A7*).

** The EuroQol-group is a group of collaborating investigators in the field of health services research and economic evaluation. Since 1987 the group is developing a standardized, non-disease-specific instrument (the EuroQol-instrument) for describing and valuing health states (for details see *A7*).

The two last remarks are on the third stage of the procedure, *the calculation of QALY's and the aggregation of quality-adjusted life years* of different branches as an aggregate measure of patient outcome following a treatment choice.

1. An important requirement for the use of the values resulting from the second stage of the procedure in QALY-calculations is the consensus between respondents on the preference of health states. Using the EuroQol-system it appeared that broad consensus existed on the preference ranking of the 13 complex health scenarios presented. Of course this number should be extended, but the degree of (international) consensus was remarkable (*A7*).
2. As we pointed out in chapter 3, aggregation is to some extent a routine statistical operation. But it may be argued that in this case additional requirements must be satisfied if the overall result is indeed to represent cardinal preference. We are not able to give a full account of this economic problem (see Harsanyi, 1955; Pliskin, 1980), but in *A6* one of the requirements seems reasonably satisfied. This observation still has to be replicated with the EuroQol-system.

Valuation of patient outcome: practical considerations

Practical considerations of the measurement procedure are to be found elsewhere (e.g. Kind, 1986; Torrance, 1986; *A6*).

In our experience the operationalization of all aspects of the procedure is demanding, whatever valuation technique is chosen. Standard gamble was been particularly difficult to convert into a feasible design; the use of visual aids in our view might change the concept of finding probability equivalents.

The use of health economists, public health officials and students fol-

lowing a course on medical decision making as respondents in many of these studies places some doubt on the generalizibility of the results[20].

Patient outcome valuation in MTA's of organ transplantation

In the American and British MTA on heart transplantation outcome valuation is lacking. Particularly in the excellent British study this lacuna is regrettable, as all the ingredients were apparently present to arrive at a cost-utility analysis. Results of the Dutch MTA's are given elsewhere (Bonsel, 1988C & 1988H; *A4*, *A6*).

5.4 Summary

Though many researchers express their belief in the existence of a yet to be developed perfect measurement instrument for health status description and a perfect measurement design for outcome valuation, none of the solutions so far in use is thought to possess all the defined characteristics. Given the cumulative effort of many experts in this field we think a perfect measurement instrument and a perfect design do not exist. However, we consider the theoretical foundations to be mature enough to justify choices on standardization among the measurement instruments and designs available.

The role of choices is different in the description of health status and the valuation of outcome. With descriptive health status, a clear predictive concept is not available and conventions refer to both the theoretical structure and its contents. With valuation of outcome the predictive concept is clear, but due to 'limitations of the real world', again we depend on conventions on how best to approximate preferences.

6 ASSESSMENT OF NEED

6.1 Introduction

The frequent use of the word 'need' in health policy suggests that there is some real but unobservable quantity, which may act as a normative value in decisions and planning activities. In section 3.6 the key position of the assessment of need within empirical MTA was indicated. Nevertheless, the scarcity of theoretical and empirical literature suggests that the elaboration, operationalization and measurement of this concept are still difficult (see Bradshaw, 1972; Williams, 1978; Holland, 1979 [p.57-]).

This chapter is therefore devoted to theoretical and practical features of the assessment of need. Section 6.2 deals with the concept of need, distinguishing between definitions based on expected health improvement and definitions based on other principles. A medical versus an economic elaboration of the first type of need definition is discussed in section 6.3, followed by an empirical approach to the assesment of need. A crude distinction between 'incidence-technologies' and 'prevalence-technologies' is proposed. Methodological recommendations are provided for four major types of technologies (preventive, diagnostic, therapeutic and rehabilitative). Section 6.6 describes our approach of need for liver transplantation. Section 6.7 presents a summary.

6.2 Conceptual choices

Although we commonly speak about the need for specific health care services, it is presumably not health care but improvement in health that patients are in need of. Health care is but a means to improve health. Even health, though valuable per se, is predominantly a condition to

arrive at other goals, like happy personal relations and welfare.

Further refinement should distinguish 'need for' from 'demand for' and 'utilization of' health care. We follow Matthew at this stage, keeping in mind that Matthew explicitly starts from the individual perspective.
'A need for medical care exists when an individual has an illness or disability for which there is an effective and acceptable treatment or cure. It can be defined either in terms of the type of illness or disability causing the need or of the treatment or facilities for treatment required to meet it. A demand for care exists when an individual considers that he has a need and wishes to receive care. Utilization occurs when an individual actually receives care. Need is not necessarily expressed as demand, and demand is not necessarily followed by utilization, while, on the other hand, there can be demand and utilization without real underlying need for the particular health service used (Mathew, 1971 in Williams, 1978)'.

Clearly the essential features of need in this quotation are the apparent scientific and ethical merits in the individual case ('... an effective and acceptable treatment or cure.'). Mathew's text (e.g. the expression 'real underlying need') suggest an objective scientific method to used.

Furthermore 'demand' is an individual function of 'need' ('... an individual considers that he has a need and wishes to receive care'). Following Mathew and others we ground need estimates for a particular health care programme or a particular medical technology on *the expected value of the health improvement* to be determined in a scientific way.

Despite the apparent attractiveness of this concept of need, other viewpoints exist. One view starts from the primacy of the judgment of the patient. In this interpretation the subjectively felt need or the manifest demand should be the starting point of an analysis of need. At the extreme this viewpoint defines the patient as a well-informed, rational, sovereign, economic agent acting on a well-priced health care market.

As put forward in chapter 3, it is exactly the invalidity of these economic assumptions in the health care market which in our view precludes a role for the individual patient in the analysis of need in empirical MTA[1].

Other views on the concept of need are visible in the debate on the pursuit of equity in health (care). These views do accept the external objective judgment instead of the judgments of patients themselves, but reject the criterion of expected health benefit. For example it has been argued that the condition of the patient, either absolute or related to some standard should determine need, apart from the improvement to be expected from medical care ("The worse off the patient is, the more he should receive help", see also section 3.8.3.). As this view denies the principle of relating ends (health) to means (health care), we think it is not acceptable as a guide to a MTA concept of need.

6.3 A medical versus an economic definition of need

Need can be conceptualized from a medical and an economic point of view[2]. The expected health improvement is essential to both. Usually the expected health improvement is a rather abstract concept: an array of outcomes may usually be encountered; certain outcomes are not equally likely and comparison between options may be difficult.

To estimate need on the individual level it may suffice to ask the patient's preference. His preference may show different principles of valuation of the relation between possible outcomes and their probabilities. Patients may accept high risks of short-term mortality and severe morbidity in exchange of a slight chance of complete recovery from a chronic relentless disease. The reverse may also be true: risk-averse patients may hesitate to subject themselves to 'safe' surgery even if, on

average, a considerable health benefit is likely. In the individual patient his preferences will be usually be accomodated (Brock, 1990).

A need estimation at the supra-individual level (e.g. the national level) requires a two-stage procedure. First the changes induced by the technology (or its alternative) to all potential beneficiaries should be empirically derived from a. the characteristics of these beneficiaries, and b. detailed knowledge about the prognosis (after treatment) related to these characteristics[3,4]. Next a uniform valuation principle should be applied (see for techniques section 5.3). The resulting values are assumed to represent the individually expected health benefits, which may lastly be compared to some criterion value.

Although the medical and the economic concept share the expected health improvement assumption, the concepts differ with respect to:

a. the nature of need as an absolute vs. a relative concept;
b. non-medical factors determining need;
c. the units in which health improvements are expressed.

Absolute versus relative concept

Doctors tend to stress the absolute character of medical need. A patient is indicated (or not) to recieve a particular treatment. From a legal and professional point of view treatment is indicated normatively the very moment a positive contribution of treatment to health is expected. This on-off definition based on the provider's judgment drives the economist Williams to speak of need as a 'quasi-supply concept' (Williams, 1978 [p.36,32]). Obviously efficiency considerations do not play a major role.

Economists tend to stress the relative character of need. Two basic assumptions are responsible for this relative concept.

First, in the economic view health care should be regarded as a commo-

dity: its consumption depends[5] on the expected benefits relative to the expected deployment of resources, and, ad libitum, to other (dis)utilities. Secondly, the desirability to apply a particular medical commodity depends on characteristics of the target technology relative to alternative options for the same resources. Theoretically the opportunities forgone by the application of a specific technology may be sought in any health care sector, across the boundary of the specific health care unit involved in the application of the target technology[6].

From the empirical perspective the operationalization of the economic view is but an extension of the medical view. However, from a conceptual point of view this 'quasi-demand' concept should be distinguished from the 'quasi-supply' concept mentioned before. The economic concept implies that a trade-off has to be made between input *and* output characteristics of the particular commodity which have to be compared with those of other commodities. The medical concept implies a judgment on output characteristics alone.

If decision-making at the societal level is to be supported - as is the rule in MTA - the economic view on need is should be preferred in the general case. Optimization of the supply of technology is usually an explicit goal of societal decision-making on medical technologies. Capacity restriction is one of the options of decision-makers (see chapter 2, the description of Phase 5 of the decision cycle). Therefore we think that efficiency considerations require MTA-information on need related to the volume and organization of the particular health-care programme. Thus the economic concept may provide the opportunity for the optimal choice of quantity and quality[7] of the health-care programme involved. Despite the theoretical preference for the economic view, we are aware that the medical view on need will suffice on many occasions, viz. if a technology shares high effectiveness with neglegible use of resources.

Non-medical factors determining need

From a medical point of view only a few factors are allowed to restrict need as defined above[8]. We can think only of the ethical or moral acceptability of a particular technology[9]. Acceptability is relevant if we try to estimate the need for e.g. abortion pharmaceuticals, neonatal units for very low birth weight babies, genetic screening facilities, etc. However, these examples of acceptability which limit the application of a particular technology, do not change the general picture: these objections function as absolute criteria. Their acceptation implies the complete absence of that particular need, their rejection the full presence.

From an economic point of view all costs and disutilities should be taken into consideration. Decreased acceptability in this view is but one type of disutility. Need depends on the balance of the input-output characteristics, the disutilities and the alternatives.

Units of expected health improvement

The last difference is related to the empirical requirements of the above-mentioned concepts. In both concepts the expected health improvement should be estimated. The operationalization, however, is different.

Medical theory does not offer a paradigmatic set of parameters to be universally used in the measurement of the medical concept of need. As the medical judgment of need depends on the benefits of the target technology alone, disease-specific parameters are a logical choice to determine need. For example, need exists if at least one expected month of life will be free of anginal pain due to the application of technology X.

Difficulties arise when the predictability of the medical benefit is low or when a wide range of outcomes is possible. Depending on the aggre-

gation principle adhered to[10], we arrive at a relatively small need only if those included have a high probability of benefit or a neglectable probability of adverse effects, or a relatively large need if all potential beneficiaries are included[11].

Problems may also arise if we have to use multiple parameters (e.g. pain and mortality). Unequivocal results (e.g. a high mortality-low morbidity pattern versus the reverse, as in some invasive versus non-invasive treatment dilemmas) may require the use of constructed multidimensional measures of benefit (e.g. QALY's).

From an economic point of view only one parameter should be used to express health improvement. Money terms would please an economist but a proxy is thought to be acceptable (Drummond, 1987). This parameter should be interpretable as (gained) 'value' (see chapters 3 and 5). Which parameter(s) and whose values (the patient, the clinical expert, or society's representatives) has already been discussed.

Economic theory provides us with various loss functions to deal with cases of low predictability of benefit at the individual level (see note 10).

Need graphs

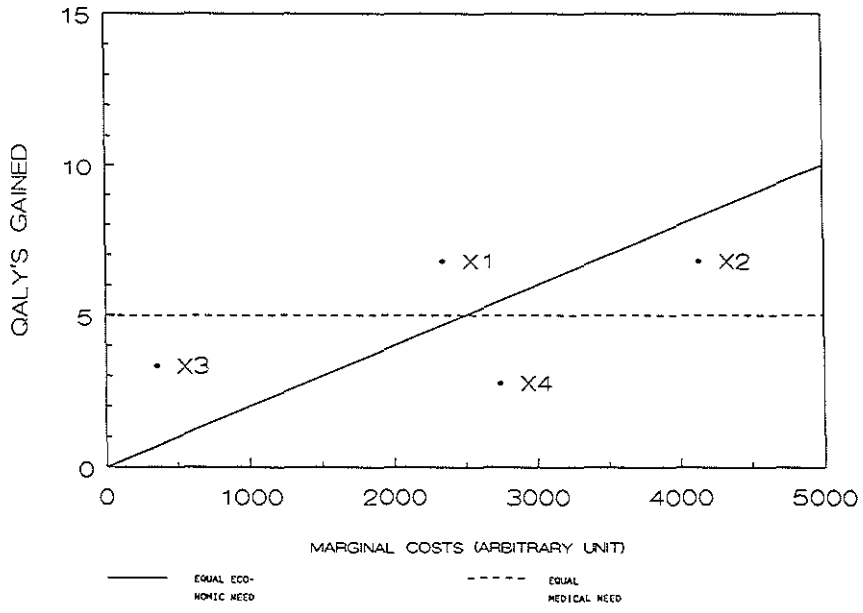
The empirical assessment of medical need investigates the epidemiology of specific medical characteristics in the population. For example, in liver transplantation data are required on the stage-specific incidence and prevalence of non-infectious decompensated cirrhosis. These data should be combined with prognostic knowledge described in chapter 4.

The empirical set-up for the economic assessment of need is not only directed at the epidemiology of these specific medical characteristics in the population. There is also 'the epidemiology' of expected resource use and disutilities. The resulting information should be combined. The

calculation of cost-utility ratios is but one method. Other aggregation rules may be adhered to (see section 3.8.1).

The results of both approaches may be presented in a need graph (see Figure 6.1). For convenience we assumed medical need and economic need to be measured in units of expected QALY's gained. Medical need[12] exists if these QALY's gained exceed a threshold (like X_1 and X_2 in Figure 6.1)[13]. Economic need exists if the QALY/cost criterion[14] is exceeded, in the example arbitrarily defined 1 QALY per 500 cost units (like X_1 and X_3). These criteria may yield conflicting results (X_2 and X_3).

Figure 6.1 Need graph describing medical versus economic need



6.4 'Incidence'-technology vs. 'prevalence'-technology

In our experience the empirical assessment of need in practice can be carried out as follows. Attention is restricted to that part of the analysis which is similar in medically and economically defined need: the epidemiology of specific medical characteristics in the population related to technology-specific health benefits.

We propose a distinction be made between an 'incidence'-technology and a 'prevalence'-technology. Incidence-technology is technology which will be applied only once in the course of the medical entity (duration independent). Except in the case of technical failures repeated application will not occur. Recurrence of the medical entity may cause repeated application but though related to the initial event it should be regarded as new medical entity. Technologies applied in high case-fatality situations (even with the technology) are a specific type of incidence-technologies: 'mortality'-technologies.

Prevalence-technology will be applied more than once in the course of a medical entity (duration dependent). Repetition may be periodical or continuous. Application is indicated by the existence of a particular ill-health state. Discontinuation is the result of death or another definite change of the health state.

Some incidence-technologies and prevalence-technologies are listed to illustrate the difference - patterns between the extremes may be observed. Examples of incidence-technologies are coronary angiography for anginal pain evaluation, heart transplantation for end-stage heart failure, physical therapy after tibial fracture treatment, vaccination for rubeola. Prevalence-technologies are for example blood sugar testing in diabetics, anti-hypertensive drug treatment, pre-eclampsial monitoring, physical therapy in cerebral palsy patients.

In our experience the choice of the most suitable empirical method to assess need depends on the above-mentioned distinction, the methods available and - of course - the available resources[15,16].

6.5 Technology-dependent recommendations

In view of the incidence-prevalence distinction, the objectives of technologies (see Table 2.2) and the characteristics of epidemiological methods, this section provides some methodological recommendations. *In all cases* we assume detailed[17] knowledge on the effectiveness of the technology involved.

In *preventive technologies*[18] need has little to do with states of actual ill health. Usually only apparently healthy people are confronted with a preventive intervention (a vaccine, a behavioural change, a dietary regimen). In most preventive strategies a large number of individuals are approached, with only a few beneficaries (who cannot be identified ex ante[19]). Estimates of need do not depend on this number of beneficaries (i.e. the number of diseased prevented) only, but also on medical and on economic considerations. Hence an empirical estimate of the need for prevention is not possible in a straight-forward epidemiological way.
A mathematical modelling approach usually is then indicated.

Diagnostic technologies may be applied in persons in apparently good health (as a screening technology) or in patients in ill health. In the latter case we assume a diagnostic process starting with a patient with complaints of an unknown disease and ending up with a description of diagnosis, health status and prognostic factors. Clearly the problems in identification of need in the former case do not differ much from those engaged in preventive technologies. However, the feasibility of an

empirical estimate grows as we walk down the diagnostic path. Diagnostic technologies applied in clearcut diagnostic entities give few problems.

Need estimation of diagnostic technologies is difficult if a. patients' profiles are difficult to define and do not relate to conventional epidemiological classifications (e.g. the need for an abdominal CT-scan), and/or if b. multipurpose application is likely to result in a multitude of heterogeneous patient profiles (e.g. the CT-scan and HBsAg-antigen testing).

Usually these difficulties prevent a straightforward epidemiological approach. Results from experimentation (preferably an RCCT in a representative setting), analysis of the health-care programme (see section 3.4) and epidemiological analysis have to be combined.

Treatment usually implies that some diagnostic label has been attached to the patient. The empirical estimate of need for a specific treatment may be relatively simple if this label is closely related to conventional epidemiological classifications. Close relation includes the absence of an additional set of conditions to be met before treatment is indicated[20]. If additional patient data are required (for example on severity of disease or contra-indications) the estimation of need is more complex. As incidence-related data outnumber prevalence-related data, need estimations on therapeutic modalities which are applied once in the course of disease are usually more easy to obtain than figures on continuous or intermittent therapies.

With respect to assessment of need, *rehabilitative technologies* resemble therapeutic technologies. As a rule rehabilitation is directed towards the improvement or the prevention of amelioration of functional status. Consequently its connection to conventional epidemiological classifications is rather loose. Health surveys and administrative data may compensate for this disadvantage.

6.6 The need for liver transplantation

The assessment of need for liver transplantation in the Netherlands was one of the issues of the empirical MTA of liver transplantation. The analysis, finished in spring 1988, is shown in detail in the appendix (*A9*; see also Bonsel, 1988G). Parallel to the assessment of the need for liver transplantation in the Netherlands, the need for heart transplantation was investigated. Although the methodology was similar, the results, which are not presented here, were less satisfactory (Bot, 1988A).

We aimed at the estimation of the need for liver transplantation following the medical concept alone. Considerations of ethical acceptability (e.g. transplantation in the very young) were neglected. At that stage the application of the economic need concept did not seem possible.
No longitudinal epidemiological studies of end-stage liver diseases and congenital liver diseases were available to allow for direct estimation.
Thus we applied three indirect methods, partially following the above-mentioned approach, defining liver transplantation as a therapeutic incidence-technology. Previous studies on the need for heart transplantation provided some additional clues (Evans, 1984; Buxton, 1985).

First the use of registry-data was investigated. We concluded only mortality statistics to be useful. The use of cancer registry data was not considered as malignancies (with only rare exceptions) were excluded from the protocol. National utilization registries (hospital admissions) could not be used due to classification problems and readmission bias.

The *mortality-based* method started from national age and sex-specific mortality data. In the case of liver transplantation the clinical syndrome (cirrhosis) fits well into a limited number of disease categories of the classification system currently in use (the ICD-9; WHO, 1977).

Adjusting factors were estimated to express the quantitative association between death due to a specific liver disease and suitability for liver transplantation. Thus we were also able to estimate the impact of future inclusion of older patients.

As a second method we used the data from the patient's registry of the liver transplantation centre (including all referrals) as a kind of utilization data. The conditions allowed for their valid use:

a. the records of all referred patients were quite detailed, particularly with regard to the suitability for liver transplantation;
b. the centre caught all patients from the Netherlands, being the only centre carrying out liver transplants;
c. geographical variations in the distribution of the prevalence of the relevant diagnoses were unlikely[21].

With this method (in the reports indicated as the *practice-based* method) current transplantation activities were extrapolated. Incomplete referral, relaxation of contraindications, inclusion of acute hepatic failure and retransplantation were accounted for.

Finally we revised available epidemiological data on patients with specific liver diseases (*literature-based* method). As these data pertained to the incidence of the diagnosis at an earlier stage of the disease, additional assumptions were necessary. Some of these data originated from other countries, particularly from the United Kingdom.

The three methods yielded corresponding results suggesting an annual need of at least 25 liver transplantations. Without changing inclusion and exclusion criteria, we estimated in 1988 a maximal need of about 70 transplantations (5 per 10^6 inhabitants) for the next, say, 5 to 10 years. These estimates were close to those of O'Grady (1988) for the United Kingdom if we applied the same selection of age-range and diagnoses to his results.

In heart transplantation we were less succesful, particularly due to the fact that clinical heart failure does not match with the ICD-9. Also the clinical information on all referred patients was less detailed and the coverage of the two Dutch programmes was less clear. Suitable epidemiological literature was even more scarce than in the case of liver transplantation. Hence we could only apply the mortality method with a low degree of certainty about its precision.

6.7 Summary

Conceptual choices and lack of data dominate the analysis of need for a particular technology. In our view any concept of need should start from expected health benefits. With regard to the empirical task, we restricted our attention to the medical concept of need. The empirical analysis of need as an economic concept still awaits elaboration.

An approach distinguishing incidence and prevalence-technologies is proposed. Even if a medical concept of need is adopted, data requirements often preclude the straightforward use of existing information.

Some favourable circumstances allowed for an estimation of the need for liver transplantation by three independent methods[22].

Despite this apparent success, this issue urgently needs further development. Improved data availability from existing national registries would contribute to this development (Vandenbroucke, 1990).

7 DISCUSSION AND RECOMMENDATIONS

7.1 Introduction

Medical technology assessment (MTA) has entered a new era (Relman, 1988; Fuchs, 1990). It has been acknowledged as an instrument which may have a beneficial influence on health-care policies. Its perspective is that of society and its objective is to pursue society's interest by supporting decisions on the health-care system.

Currently at least three major types of MTA can be distinguished: meta-analysis, the Delphi technique and empirical MTA. This thesis focusses on the latter type of information generation which has shown to be a demanding research activity. It calls on the inventivity of the researchers to overcome methodological complexities and limitations of data acquirement. It calls on the cooperation of health care providers without offering them prior certainty about the nature of the results and conclusions. Lastly the valid use of it results requires a politically well-structured decision-process.

Some have doubted the relevance of empirical MTA-research activities and have questioned the additional value of empirical MTA above clinical research. In their view the current quantitative growth of MTA-like research should be interpreted as a temporary political affair. This thesis defends the opposite position.

This chapter will subsequently answer the three questions put forward in section 1.7.:

1. What is the scientific basis of an empirical MTA?
2. Which conditions determine the fruitful application of empirical MTA results?
3. How can MTA be developed further?

Most attention will be given to the first question about the scientific basis of MTA, to which the first part of this thesis was devoted. Section 7.2 gives an annotated account of the methodological features of empirical MTA-research.

Availability of a sound methodological approach is but one condition to ensure suitable results of MTA-research. The experience from the Dutch transplantation MTA's, particularly the MTA of liver transplantation, provided some conclusions on how to arrive at results which are both scientifically acceptable and politically useful. These conclusions are presented in section 7.3.

Section 7.4 casts an eye to the future. The validity of results and their subsequent use is only one condition for further development of this branch of health sciences. We present an outlook on the additional work to be done to settle MTA within the family of health sciences.
The chapter concludes with an epilogue.

7.2 The scientific basis of empirical MTA

The purpose of empirical MTA is the prospective evaluation of a health-care programme from the true society and true time perspective, comparing it to the available alternatives on the level of aggregation corresponding to that of the decision-maker.
The legitimacy of the claim of a specific MTA-methodology is supported by the existence of a number of core issues in MTA which, if dealt with in a specific way, indeed make up such an evaluation. These core issues are: the action of the technology, the effects on patient outcome, the effects on the use of resources, the need for the technology and the role of actual and future uncertainties. Particular circumstances may justify

the addition of more issues but from the MTA-literature we conclude that these issues are *complementary* to the topics on the effects and the future impact of the technology. The general methodological approach to these issues is described analytically in Table 7.1. This approach may to some extent be regarded as an extension of the economic evaluation concept as described by Drummond (1987).

Table 7.1 Methodological approach of an MTA from a societal perspective

a. a dynamic disease model linking the occurence and the development of disease to survival, health status and other events amenable to intervention;
b. a technology application model, linking the employment of resources to the disease model via defined interventions;
c. an overall structure of the model which allows for comparison over time of alternatives;
d. uniform concepts for input, throughput, output and need, which are interchangeable in all branches of the model;
e. uniform disease-non-specific measurement instruments for output measurement; uniform setting-non-specific measurement instruments for input measurement;
f. a comprehensive approach to uncertainties.

This approach has generally been followed in the Dutch transplantation MTA's and its predecessors in the United States and the United Kingdom (Evans, 1984; Buxton, 1985).

Methodology to analyze the issues separately as a rule already exists (see chapter 3). Survival analysis may be found in biostatistics, analysis

of health status in clinical psychology and psychometrics, cost analysis in economics and the assessment of need in epidemiology. The point of distinction in Table 7.1 is the multidisciplinary, quantitative approach.
All efforts are directed towards an integrated comparison of the consequences of choices. Commensurability of concepts and their operationalizations is therefore mandatory. As the choices are made from society's point of view, comparing benefits and burdens, the influence of economic concepts and analytical structure in MTA may be explained. Empirical MTA thus represents a scientific basis for societal choice (Farber, 1982).

A major alternative to empirical MTA is the informal and/or formal consultation of experts prior to decision-making. Perhaps another is the renouncement of specific information generation prior to judgment, as removing the veil of ignorance may be costly. Here we shall discuss only the first option.

Some weaknesses are inherent to expert consultation.
First, qualitative and quantitative knowledge on alternative treatment patterns in high-technology settings is limited. Secondly, there is usually limited expertise on costs and need compared with knowledge about the effects on patient's outcome. Thirdly, it appears to be difficult to express opinions on patient's outcome in an intersubjectively quantitative way; quantification of opinions on the other issues is even more complicated and few experts are capable of making complex integrating forecasts. Fourthly, the interests of the consulted experts may blur fair judgment. Hence, though a discussion of the results of empirical MTA with experts in the field is indispensible, expert opinion may not easily replace empirical MTA-investigations (see also chapter 2).

Some specific methodological recommendations for empirical MTA are given next.

In *description of the action of the technology* we advocate a distinction between qualitative health-care programme description and quantitative analysis of programme dynamics. The former should pay due attention to existing alternatives, including those within the non-teaching hospital setting. In the quantitative part we suggest the routine application of survival analytical techniques.

Experience from the Dutch transplantation studies may support this recommendation. In the liver transplantation study it was thus possible to compare the Dutch programme with the average European experience (Bonsel, 1988E; *A1*). In the heart transplantation case the analysis was used to predict organizational and outcome consequences of various strategies of patient selection (Van Hout, 1990).

Available evidence (e.g. NIH, 1983; Evans, 1984; Buxton, 1985; more generally, the issues of the Journal of Technology Asessment in Health Care) suggests that in current MTA-practice this issue usually lacks elaboration. From our experience we conclude that a clearcut explicit organization of patient flow in the clinical pilot programme with an administrative account of decisions is essential for success.

Analysis of survival and health status is the cornerstone of the analysis of patient's outcome. It may be carried out from a medical-descriptive and from an economic-valuing point of view. In survival analysis the consequences of differences between these views are limited, in health status measurement they are considerable.

In *survival analysis* we advocate the thorough use of the available statistical methodology, which is equally useful from both views. Cox' proportional hazards model (1972) in particular has opened wide perspectives enabling, for example, the analysis of time-dependent effects of predictor variables and the estimation of survival consequences of new technologies in the absence of information from randomized controlled

clinical trials (RCCT's).
The Cox' model was used for several purposes in the Dutch transplantation MTA's (*A4* and Van Hout, 1990).
Available evidence on survival analysis has shown that the major problems are technical mistakes (see Bonsel, 1988F for an account of liver transplantation literature) and the slow adaptation of powerful statistical innovations. On the design level the impopularity of the RCCT in clinical practice is more important. Clearly these problems are not specific for MTA and their solution should be found in a change in clinical attitude and an active responsibility of MTA commissioners for design aspects.

In *descriptive analysis of health status* we suggest the use of a general model which allows for various operationalizations. The role of the time horizon should be explicitly stated, as it directly influences the phrasing of items and the scheduling of interviews. Depending on the precise purpose of the description and the style of operationalization chosen, existing measurement instruments may be chosen or new ones developed.
An extensive set of criteria for the choice of measurement instruments suitable in health status research is available (chapter 5; Bot, 1991).
In the transplantation studies we arrived at a health status model, derived from Campbell and others (1973; chapter 5; Bonsel, 1989A [p.115-133]). The choice of suitable measurement instruments and of an adequate timing of the surveys was a compromise resulting from theoretical, practical and financial constraints. The choice for computer assisted interviewing of patients turned out to be a lucky one, contributing to high efficiency of the enterprise. Statistical analysis was difficult due to right and left censoring of data.
From our experience we conclude the major conceptual problems in descriptive health status analysis to be the infrequent use of an explicit general model, the lack of well-documented questionnaires, the lack of

acknowledgement of the disease-non-specific perspective in MTA and the conflicts of interests which may easily arise in this issue. Only the latter two problems specifically pertain to MTA. Standard methodology is available, but incentives are lacking to apply already available knowledge from a shared perspective.
A separate issue is the frequent absence of the experimental design which is equally detrimental to this issue as to survival.

Economic analysis of outcome usually takes the form of a cost-utility-analysis (chapters 3, 5). Up to now the ideal approach, cost-benefit analysis, has turned out to escape acceptable operationalization. Cost-utility analysis requires the separate outcome parameters (survival and health status) to be expressed in one unit. The resulting unit represents the 'utility' of outcome. Numbers attached to utility should, like money, reflect 'value'[1], hence we speak of valuation in stead of description. We advocate the use of so-called Quality Adjusted Life Years as the best technique now available to arrive at utilities. The quality adjustment may be arrived at by a three stage procedure (see chapter 5). The use of QALY's or any other measure of utility rests on many assumptions, which require for carefully designed measured instruments.
Within the context of the heart transplantation study we used an instrument for quality adjustment which was particularly suitable for transplantation patients (*A6*). Later we developed such an instrument for international use together with a group of European colleagues (*A7, A8*).
This issue, and in particular the use of QALY's, is perhaps the most controversial in MTA. Some of the controversy is on the use of measures of effectiveness in general, as these are not invariant to age, gender, race etc. Both from a philosophical and a pragmatic view we do not see any valid alternative: any one-unit estimate will show preference for those with the largest potential of loss of healthy years: the severely

diseased young with a condition amenable to intervention. If we accept the use of effectiveness measures as a reflection of the value of the output from the deployment of resources, up to now no alternative to QALY's exists: healthy-years-equivalents are only an acronym attached to a specific variant of QALY's (Richardson, 1990).

In reply to another part of the criticism, it should be stated that on the population level broad consensus appears to exist on the relative desirability of multidimensionally described health states with the same duration (*A6, A7*); agreed this is only one piece of the puzzle, but it supports the hypothesis that judgment on the (un)desirability of complex health states *is* possible.

Recommendations are directed at a wider application of a standard 'QALY-toolkit' (the term is from Paul Kind). Choosing one standard may appear to be more important than lengthy discussions about which standard is best.

Cost analysis is rather new. The principles formulated by Drummond may serve as a first guide. We suggest additional attention should be given to costs of care as opposed to costs of research, to the expected impact of scale effects, to the relevance of patient costs, and to the issue of future medical costs. Admittedly in the Dutch MTA's attention to most of these topics has been limited: the empirical task of description of direct costs was in itself demanding.

However, the tendency of making gross comparisons of the cost-effectiveness of technologies makes some of these topics urgent. For the same purpose reference values for broad and small disease categories are urgently needed (for an example see Koopmanschap, 1991).

Additionally the negative consequences of absence of the RCCT-data to cost-analysis should be mentioned. It may be even more difficult than in outcome analysis to acquire next-best estimates.

With regard to *the determination of need* the expected benefit from a particular technology in our view should determine the degree to which the need for a technology exists. This methodological principle connects the methodology of need to the methodology of the other core issues (see chapter 6) and to the optional issue of distributional aspects.

Some people have suggested that 'subjective' need (more precisely: manifest demand) be accepted as a basis for need estimates. As altruism is uncommon, it is unlikely that this approach will yield efficient nor equitable results. An economic interpretation of need additionally requires specific cost information.

In liver transplantation medical need could be estimated with acceptable certainty due to some favourable circumstances (chapter 6). In heart transplantation this was not possible due to lack of epidemiological data on heart failure (Bot, 1988A). In neither case did we try to arrive at the economically defined need.

We conclude so far that two major problems exist in the determination of need: lack of conceptual clarity and severe underestimation of the empirical task. However, the alternative of asking experts incorporates serious disadvantages as we experienced in the liver transplantation study. Thus we recommend the institution of a feasibility study on the routine use of the national hospital admission registry for this question.

We regard *the analysis of uncertainty* (scenario analysis) as a separate issue. The nature of uncertainties differs considerably but their technical treatment may be similar (e.g. by stochastic modelling).

The liver transplantation and the heart transplantation study showed the relevance of thorough scenario-analysis.

The major problems in scenario-analysis are the technical and political requirements, and the difficulty of presentation. All parties in MTA would benefit from a shared participation in this part of the analysis.

Finally, we provide some remarks on two issues regarded as optional. In chapter 3 it is argued that *equity considerations* have sofar been of limited relevance in MTA. Data requirements are extremely high (particularly in the elaboration of Rawl's equity rule). Even in the case of liver transplantation we do not expect equity analysis to be possible; moreover the practical importance may be doubted.
Legal and ethical issues should not be neglected but we think they are important *after* we have factual knowledge on the abovementioned isues.

Summarizing section 7.2 we think the question of whether there is a specific scientific basis for MTA has to be answered affirmatively.

7.3 Conditions which determine the fruitful application of empirical MTA-results

The answer to this is twofold. The first part describes the political conditions, the second the investigational conditions.

The *political environment* should generate relevant questions and should employ explicitly the information supplied. We recommend a decision-cycle be instituted as proposed in chapter 2.
In the Netherlands the National Health Insurance Board proposed a decision-cycles of this kind for all new technologies and the decision-process on heart and liver transplantation may be regarded as an exercise in this respect. Despite this favourable development in the Netherlands and similar developments abroad, some features of the current political context of MTA could be improved. They are presented in order of the phases in the decision-cycle.

As resources are limited, a *rational selection of technologies* to be subjected to MTA is required.

Criteria for selection are suggested in chapter 2 and reflect the potential impact of the MTA on health and health care. This is to some extent contrary to current policy to institute MTA's for ad hoc budgetary or ethical reasons alone.

MTA requires a *clear definition of the decision problem*. Relevant aspects should be mentioned, criteria be stated, where necessary.
From our experience we conclude that MTA's often lack a prior description of the study objective. Sloppiness in this respect can only partially be compensated for afterwards by retrospective tuning of the results.

Given a well-defined decision problem, the most *suitable method of information generation* should be chosen. As all additional research activities usually imply considerable investments, the available literature should be investigated first. A checklist of criteria for judgment is proposed in chapter 2.
We conclude that the major drawbacks of existing information are lack of relevance due to cultural barriers (particularly between the Netherlands and the U.S.) and partial coverage of the core issues. However, the price of suboptimal suitability of existing information should be balanced against the costs of additional investigations and against the disadvantages of delaying the decision.

The recommendations on empirical MTA, one of the prototypes of information generation, were presented above and will not repeated here

Some considerations on the subsequent *political valuation of the collected information, the subsequent decision-making and the implementation of decisions* are given next. Quick use of the MTA-results in the decision-making should be attempted. Political inertia after a MTA may easily place doubt on the usefulness of MTA's. A clear relation between considerations and valuations on the one hand, and the supplied infor-

mation on the other will contribute to the credibility of the procedure. Explicit relations between MTA and decisions were quite obvious in the Dutch transplantation cases (Van Rossum, 1990). Decision-making in these cases started during the MTA, though formal decisions were made afterwards. The link between the final MTA-results and the decision in our view was on average clear.
In current MTA-programmes we recommend that more attention be paid to the anticipation of political decisions.

Lastly we assume *a feedback mechanism* usually to be relevant in the phase after decisions have been made. Experience with monitoring and repeated evaluation is limited. It has recently been decided to monitor medical technologies regulated by the Hospital Provision Act (Art. 18). Heart transplantation and probably liver transplantation will be covered by this decision in the future.
We expect that the withdrawal of proven dysfunctional technologies and particularly of technologies judged to be too expensive will be difficult. Irrespective of the status of the technology (accepted, rejected or something between) the proposed feedback mechanism requires a vivid political environment and a receptive attitude of clinical centres involved for the required organization of data collection.

The suitability of MTA-results is not only related to the existence of an appropriate political context. Empirical MTA's requires additional *conditions* for investigation to be met. We list five of them.

First, the MTA-research team should be *multidisciplinary*, sharing medical, economical and mathematical expertise.

Secondly, *standardization* of concepts, measurement instruments and integration techniques is required to enable comparitive judgment.
With standardization we enhance both technical and political efficiency.

A third condition is a good *match of the clinical pilot programme(s)*

to the empirical MTA. This frequently requires a particular set-up of the clinical programmes, inducing some change in the particular clinical research programmes concerned. In the liver transplantation study some questions were difficult to adress due to partial misfit of the clinical pilot programme to the study objectives.

A fourth related issue is the *independence* of MTA-investigators. Empirical MTA's with potential high impact should preferably be executed by independent investigators. Otherwise the magnitude of the interests will otherwise create serious risks of bias. Independence also implies prior arrangements on scientific publication of results.

Lastly during the MTA the *political position* of the parties should be kept clear. Policy research like MTA may exert considerable impact on the object to be studied and the reverse may also be true. This mutual influence is inevitable and may even be fruitful. To the extent the information exchange between the medical community and health-care authorities is stimulated, this intermediate role of the MTA-investigator is justified.

From this section we conclude that conditions may be defined which enable fruitful application of empirical MTA results. In the Dutch transplantation studies these conditions were generally satisified.

7.4 Conditions which influence the future development of MTA

We are aware that the answer to the third question posed in section 1.7 on the further development of MTA implicitly supports the idea of independent analysis of medical technologies from a societal viewpoint. Our answer on the long-term viability of MTA departs from desirable

scientific and policital developments and inevitably includes some speculation. Evidence will be drawn from experience in the Netherlands and from developments abroad, particularly in the United States, where MTA's were already being carried out in the late seventies.

On the *scientific level* the relation of MTA methodology to clinical research and clinical practice should be developed further.

In the pessimistic scenario MTA methods are regarded as a set of tools to be included in clinical research to some convenient degree. Its inclusion is only to satisfy sponsor's and scientific publisher's demands. In the optimistic scenario the clinicians recognize the long-term importance of a 'dialectical' relation with MTA-investigators and their methods. A cooperative attitude is also needed for the theoretical and methodological challenges in MTA[2]. As empirical MTA's frequently provide opportunities for additional research, clinicians and third parties alike should be prepared to support these activities[3].

We are observing a growing demand from companies and health care authorities for empirical MTA, as piggyback research on already existing clinical investigations. Unaltered, the design of these investigations and the predefined methods of data collection and analysis give little opportunity for empirical MTA satisfying the criteria mentioned earlier. Hence we recommend prior involvement of MTA-investigators in the design phase of these studies.

The scientific relation to the methodological neighbours, clinical epidemiology, biostatistics, clinical decision making, health economics, psychometrics, and operations research should also be established.

The viability in our view also depends on the self-supporting capacity of MTA units. The time involved in creating experienced multidisciplinary teams asks for a consistent grant policy at the national level. We expect most benefits from the support of units of at least moderate size.

As a last recommendation we suggest that the university management adapts to the organizational demands of this type of research. In the sector of health care information the environment changes rapidly. Flexibility is needed to keep up with the developments as these may be observed in the hospital and in non-university research settings.

On the *political level* MTA should be increasingly regarded as an obvious cost-effective response to particular decision problems (see chapter 2)[4]. The utility of being uninformed is well-known, particularly in the health-care market. Hence future developments are difficult to predict. However, we expect that both health insurance companies and consumer organizations[5] will increasingly push health care providers to demonstrate the costs and effects of health services supplied.

At least three scenarios may be thought of (a mix is likely).

1. an increase of clinical centres incorporating for example a health economist on a structural basis to comfort the external demands[6].
2. an increase of trial administration and business administration offices, offering information technology on a WYW|WYG (what you want is what you get) base; information generated thus usually will not parallel information from MTA's.
3. a steady growth of independent MTA-research despite some pressure from the health-care market forces and general anti-regulatory policies[7].

In the end we expect that the attitude of health insurance companies and consumer organizations will be decisive. When these parties recognize the adverse effects of a free market of health care[8] and when they accept that the value of technologies is only loosely related to consumer satisfaction, the political position of MTA will be established.

7.5 Epilogue

The current technological explosion in medicine has provoked many reactions. Sceptical reactions from those judging the scientific advance from the net effects on the population's health[9]. At the extremes both averse and supportive reactions may be observed.
In this thesis medical technologies are considered a priori as morally neutral. Critical analysis - MTA - and rational judgment are assumed to be possible and to address the ultimate value of medical technologies from a societal standpoint of view. This critical position provides us with a scientific basis for MTA.

The assumptions of the quantification of medical benefits and societal efforts, and the structure of the proposed analytical approach reflect an economic interpretation of health care. In our view this interpretation does not originate from the assumed budgetary problems in health care giving rise to MTA-activities. Moreover the economic interpretation is a consequence of a general change of the societal perception of medicine. Public awareness of the scarcity of resources parallels the emancipation of the patient towards an assertive economic agent, the conscious exploration of his agency function by the health care provider and the unprecedented competition between health care organizations. This metamorphosis from religion to trade implies a change towards a more symmetrical relation of patient and provider and ultimately provides us with a future political basis for MTA.

SUMMARY

This thesis concerns a recently developed method of analyzing technologically induced changes in medical care. This method is known as 'medical technology assessment'. Essentially its application should yield information to support decision-makers in the health care system.
The results of our experience in the development and application of methods for medical technology assessment are presented in two parts. The first part gives a general theoretical overview of the subject. The second part is added as an appendix. It embodies the empirical part of the thesis and demonstrates the results from practical applications of the proposed methodology. This summary is restricted to the first part.

Chapter 1 starts with a general *description of changes in health care and the contribution of new technology* to these changes. The systematic search for information on new technology is explained as a natural reaction of all actors in health care. A growing awareness of the economic relevance of this information is obvious.

Health-care authorities have gradually exchanged their passive and permissive role for a more normatively active role. The emancipatory development of consumer organizations in health-care is remarkable too. Both changes breach the traditional monopoly of health-care providers. The introduction and diffusion of new medical technology are no longer the domain of professional interest alone.

This thesis draws most of its illustrative evidence from the Dutch MTA of liver transplantation, a study instituted by Dutch health-care authorities (1985-1988). The history of liver transplantation in the Netherlands and the position of this MTA is summarized.

Three key questions are formulated which are considered to be relevant in the judgment of current and future value of MTA:

a. what is the scientific basis of an empirical MTA?
b. which conditions determine the fruitful application of empirical MTA results?
c. how can MTA be developed further?

Chapter 2 presents a prototypical *description of a rational political response to technologically induced changes* in the health-care system. This description apparently fits current health policies in the Netherlands which aim at information-supported societal decision-making on new medical technology with high impact. The response is analytically described as a *decision-cycle* with six phases: a. recognition and selection of technology b. definition of the decision problem, c. definition of the information need and selection of the information generating modality, d. information generation, e. decision-making and implementation, and, if appropriate, f. monitoring and repeated assessment. The roles of health-care authorities, MTA-investigators, and clinical experts vary according to the phase.

This thesis focuses on the phase of information generation. Three information-generating modalities to be used in this phase are mentioned, viz. meta-analysis, Delphi-study, empirical MTA. The remaining chapters are devoted to the methodology of empirical MTA.

Chapter 3 presents a conceptual *description of the methodology of empirical MTA*[•]. A standard format is proposed consisting of five core issues: the action of the technology, the effects on patient outcome, the effects on the use of resources, the need for the technology and actual and future uncertainties.

• Empirical MTA is further usually indicated by MTA sec.

Methodological standardization is advocated, particularly the use of RCCT's and the synthesis of results by means of quantitative models. The arguments for standardization are, in our view, compelling: technically it enables medical and economic analysis to be combined with forecasting, politically it enables the comparison of various MTA-results. Some conceptual remarks are presented on the five core issues.

The *description of the action of a technology* consists of a qualitative and a quantitative part. The qualitative part gives full account of the new technology and existing alternative interventions alike. The quantitative part applies flow chart and survival techniques to describe quantitatively the course of disease and patient's history over time.

Conceptually the *effects on patient outcome* may be described from a medical and an economic point of view. In the first case survival and health status are separate concepts. In the latter case we usually use life years gained (with or without quality adjustment) as a comprehensive unit of outcome.

The *assessment of costs* includes an estimation of both direct and indirect costs. The rationale of inclusion of future medical costs is discussed. The distinction between costs for research and costs for care is advocated, particularly if cost estimates of a clinical pilot programme serve as a decision-making, planning or budgeting tool.

As with patient outcome we distinguish a medical and an economic approach to the *need for a technology*. In both cases the need as perceived by the individual is regarded not to be a reliable and valid instrument for the purposes of MTA, particularly not with new technologies. The quantity of expected health benefits to be determined objectively forms the basis for both approaches.

The *analysis of uncertainties* distinguishes between actual and future uncertainties. Actual uncertainties are of various natures: they consist of

unknowns due to in complete information, 'intangibles' (measurement problem) and 'imaginables' (definition problem). Future developments are uncertain by definition; usually they are to some extent predictable as an autonomous trend. Additional policy analysis provides insight into uncertainties due to changes in the environment of the technology. The quantitative approach of all uncertainties has much in common and rests upon the previously mentioned quantitative model.
Two technical features of the proposed quantitative empirical approach are discussed separately: formal synthesis and subgroup analysis. *Aggregational techniques* within MTA has received some criticism. We justify their application as an ubiquitous statistical technique. Political arguments could be added. The political implications of *subgroup analysis*, a second standard statistical technique, are demonstrated. The last section in chapter 3 is devoted to distributional, legal and ethical considerations.

Chapters 4 to 6 provide backgrounds to survival analysis, health status analysis and the assessment of need. To prevent duplication emphasis is placed on methodological backgrounds supplemented with an abstract from our experience, shown in more detail in the appendix.

Chapter 4 describes *survival analysis* in MTA. It is argued that survival analysis predominantly consists of the enforced application of already available statistical tools. Supplemental calculations are necessary to arrive at life years gained as an economic unit of effectiveness. The challenge of many MTA's is the estimation of survival consequences of an intervention in the absence of any information from randomized controlled clinical trials (RCCT's); this absence is often observed in new technology. Several approaches to the estimation of control-group survival ('shadow survival') in this situation are presented, including the use of prognostic models. The technical approach of the last method

which allows for quantitative prediction is shown. These predictions may be used both at the individual and at the aggregate level. Also the computation of the number of life years gained due to the target technology is possible, provided the censored observations are treated appropriately.

From our experience in liver transplantation we conclude that the estimation of shadow-survival is possible if:

a. quasi-experimental control group data are available, or

b. intervention delay data (e.g. waiting list data) are available, or

c. historical control group data are available.

Additionally prognostic modelling is usually feasible if individual data from a., b. or c. are available. In all cases homogeneity of reference and index patients with regard to diagnosis and prospect should be sufficient.

Chapter 5 presents some methodological considerations on health status measurement both from a descriptive and a valuing point of view.

Starting with descriptive disease-non-specific health status, it is argued that health status is an unobservable theoretical concept which is similar to many clinical concepts. For validity to be established, this state of affairs requires a. consensus on a theoretical structure containing the particular concept and b. consensus on facts to be predicted from a priori specified parallel or related concepts.

For *descriptive purposes* a general structure for the health status concept is proposed. If, subsequently, a questionnaire is developed, the choices of the exact elements of the structure (e.g. the dimensions) and of items and answer modalities, may be guided by criteria of either statistical efficiency or of interpretative soundness, each having their own virtue. Additionally a comprehensive set of criteria is available to judge the feasibility and quality of descriptive health status measurement instruments.

Within health status measurement 'time' occupies an important position: the period to which questions on health status refer and the interview frequency are closely related and should reflect the underlying concept of health status.

Our experience with health status description in the Dutch transplantation MTA's is described. Two features of the descriptive analysis appear to apply to a general problem in many MTA's: the difficult statistical analysis of right and left censored data and - as with survival - the difficulty of comparitive analysis in the absence of a control group.

The second part of the chapter is devoted to health status measurement as part of the *economic measurement of patient outcome*. Up to now the best available technique to arrive at an economic value of patient outcome is the calculation of life years gained or, if possible, of Quality Adjusted Life Years (QALY's) gained. QALY's require subsequent measurement of the patient's health status, construction of health status descriptions (vignettes, scenarios) and subsequently valuation of these descriptions. The use of QALY's in so-called cost-utility analyses rests on many assumptions, which ask for carefully designed measured instruments. As a result of a collaborative study we were able to develop an instrument for health status valuation in this context (EuroQol), which is internationally available. With this tool the validity of one critical assumption of QALY calculations could be confirmed: the interpersonal consensus about the preference order of complex health status situations with a fixed duration.

The results of cost-utility-analysis in the Dutch transplantation MTA's support the feasibility of the QALY procedure.

As a separate issue chapter 5 reports the rewarding experience with personal computer-assisted interviewing in health status measurement.

Chapter 6 is devoted to *the assessment of need* for a particular technology. Conceptual choices and lack of data dominate. First a definition based on expected health improvement or one of the available definitions based on other principles must be chosen. These conceptual choices are the same as those encountered in the definition of equity (see the end of chapter 3). In our view any concept of need should start from the first definition. Next, a medical and an economic elaboration of this need definition is discussed. An empirical approach is offered, distinguishing between 'incidence-technologies' and 'prevalence-technologies'. Consequently methodological recommendations are provided for four major types of technology (preventive, diagnostic, therapeutic and rehabilitative). Favourable circumstances allowed estimation of the medically defined need for liver transplantation.

Chapter 7 adresses the three questions mentioned in the introduction.

What is the scientific basis of an empirical MTA?

From the previous account it will be clear that this thesis defends the position this basis consists of a number of core issues, which are approached specifically. Rigorous standardization, quantitative modelling and integration of medical and economic concepts are the ingredients. In many issues standard tools are available or, even more important, standard methodology exists to arrive at these standard tools. The existence of methodological choices is not denied, but in our view these choices reflect both a vivid multidisciplinary debate and a close relation between scientific labour and societal discourse.

Which conditions determine the fruitful application of MTA results?

The application of empirical MTA results apparently depends on the satisfaction of political and investigational conditions. With regard to the former a rational and explicit decision-process will clearly add to an effective and acceptable use of MTA-information. In the Dutch case the National Health Insurance Board (Ziekenfondraad) has proposed a decision-cycle as described in chapter 2. Though not entirely operational, the proposal has already resulted in three pilot MTA's (of in vitro fertilization, heart and liver transplantation) and subsequently positively influenced decision-making on these and other new technologies.
Second to this are conditions for investigation, including the existence of independent multidisciplinary MTA institutes, the political support of these institutes regarding the application of standard methodology, and a match of the clinical pilot programmes to the MTA requirements.

How can MTA be developed further?

The answer to the last question on the future development of MTA is obviously of a political nature. The previously mentioned conditions will of course influence its development but we do not think they are essential. Moreover we think two other, general developments will be decisive. First, the development of scientific relations between MTA institutes and both their methodological and clinical neighbours. Secondly, the development of a political momentum which acknowledges the permanent necessity of MTA research, as a natural response to asymmetrical economic relations in the provision of medical care. We expect this to depend on the nature of the emancipation of the medical consumer.

SAMENVATTING

Dit proefschrift beschrijft een methode om veranderingen in de gezondheidszorg onder invloed van medische technologieën te analyseren. Zij wordt wel aangeduid met 'medical technology assessment' (MTA)*.
Het uiteindelijke doel van MTA is het genereren van informatie ter ondersteuning van besluitvorming in de gezondheidzorg. Onze ervaring in het ontwikkelen en toepassen van methoden voor MTA worden in dit proefschrift beschreven. Het proefschrift bestaat uit twee delen.
Het eerste deel geeft een algemeen theoretisch overzicht over MTA. Het tweede deel is toegevoegd als een appendix en bevat empirisch toepassingen van de in het eerste deel voorgestelde methodologie. Hier wordt alleen dit eerste deel samengevat.

Hoofdstuk 1 beschrijft de effecten van *veranderingen in de gezondheidszorg met speciale aandacht voor de bijdrage van technologische vernieuwing* aan deze veranderingen. De structurele verzameling van informatie over nieuwe technologieën is een verklaarbare reactie van alle actoren in het veld van de gezondheidszorg. Zij zijn zich in toenemende mate bewust van het economisch belang van deze informatie.
De rol van de gezondheidszorgbestuurders (in Nederland kan men denken aan het Ministerie van WVC, de samenwerkende ziektekostenverzekeraars, de Ziekenfondsraad, het Centraal Orgaan Tarieven in de Gezondheidszorg, de Gezondheidsraad en diverse koepelorganisaties) is aan verandering onderhevig. Een passieve en afwachtende rol wordt in toenemende mate ingewisseld voor een actieve medebepalende rol. De emancipatie van consumentenorganisaties op dit terrein is evenzeer opmerkelijk. Het traditionele monopolie van de zorgaanbieders wordt daarmee doorbroken.

---

* Een bevredigende Nederlandse vertaling ontbreekt.

De introductie en verspreiding van nieuwe medische technologieën zijn daarmee niet langer een zaak van de professie alleen.
Bij de beschrijving van MTA put dit proefschrift voor de voorbeelden in hoofdzaak uit de Nederlandse MTA van levertransplantatie, een studie die op initiatief van de Ziekenfondsraad en de Gezondheidsraad werd uitgevoerd (1985-1988). Om die reden wordt de geschiedenis van levertransplantatie in Nederland en de plaats van deze MTA kort beschreven.
Hoofdstuk 1 besluit met drie vragen, die van toepassing lijken voor een beoordeling van de huidige en toekomstige waarde van MTA:

a. wat is de wetenschappelijke basis van MTA?
b. welke omstandigheden begunstigen het gebruik van MTA-resultaten?
c. hoe kan MTA in de toekomst verder worden ontwikkeld?

Hoofdstuk 2 geeft een *modelbeschrijving van een rationeel-politieke aanpak van technologische veranderingen* in de gezondheidszorg. Deze beschrijving blijkt van toepassing in Nederland. Hier te lande wordt gestreefd naar maatschappelijke besluitvorming over invloedrijke nieuwe technologieën op basis van zakelijke informatie. Deze aanpak kan in feite beschouwd worden als een *besluitvormingscyclus,* waarin zes fasen zijn te onderscheiden: a. herkenning en selectie van de te beoordelen technologie, b. definitie van het beslissingsprobleem, c. definitie van de informatiebehoefte en van de informatieverzamelingsmethode, d. informatieverzameling, e. evaluatie, besluitvorming en implementatie, en tenslotte, indien opportuun, f. voortgezette controle en re-evaluatie.
In ieder van deze fasen spelen bestuurders, MTA-onderzoekers en de toepassers van de technologie een eigen rol.
Dit proefschrift handelt over de vierde fase, de informatieverzameling. Tenminste drie typen worden onderscheiden t.w. meta-analyse, Delphi-onderzoek en MTA primair op basis van empirisch onderzoek.

De resterende hoofdstukken zijn gewijd aan de methodologie van dit laatste type.

Hoofdstuk 3 beschrijft met name de *conceptuele aspecten van MTA-methoden*. Een standaard structuur wordt voorgesteld bestaande uit vijf obligate en twee facultatieve onderwerpen. De vijf sleutelonderwerpen zijn: de werking van de technologie, de effecten op de patiënt, de effecten op het middelenbeslag, de behoefte aan de technologie en huidige en toekomstige onzekerheden.
Methodologische standaardisatie wordt aanbevolen, met name de toepassing van RCCT's en de synthese van uitkomsten door middel van kwantitatieve modellen. Er zijn dwingende redenen voor standaardisatie: in technische zin is het een voorwaarde voor het combineren van de diverse onderdelen tot een toekomstvoorspellend model, in politiek zin is het een voorwaarde voor vergelijkbaarheid van MTA's.

Hierna volgen enige conceptuele aandachtspunten ten aanzien van de voorgestelde vijf sleutelonderwerpen van een MTA.
De beschrijving van de *werking van de technologie* bestaat uit een kwalitatief en een kwantitatief deel. Het kwalitatieve deel beschrift de nieuwe technologie en haar alternatieven in vergelijkbaar detail. Het kwantitatieve deel past zgn. flow chart en overlevingstechnieken toe om het ziektebeloop en de gang van de patiënt kwantitatief te beschrijven.
De *gevolgen voor de patiënt* kunnen op twee elkaar aanvullende manieren worden beschreven, namelijk vanuit medisch en economisch gezichtspunt. In het eerste geval onderscheidt men overleving en ziekte-aspecifieke kwaliteit van leven (gezondheidstoestand) als twee gelijkwaardige uitkomstmaten. In het tweede geval worden levensjaren, eventueel met kwaliteitscorrectie, als alomvattende uitkomstmaat gebruikt.
Voor het vast stellen van het *beslag op maatschappelijke middelen*, kort-

weg de kosten, wordt onderscheid gemaakt tussen directe en indirecte kosten. Beargumenteerd wordt dat alle toekomstige directe kosten moeten worden inbegrepen. Het onderscheid tussen kosten voor zorg en kosten voor onderzoek verdient aandacht, vooral als kostenschattingen van een klinisch ontwikkelingsprogramma worden gebruikt als hulpmiddel voor besluitvorming, planning of financiering.

Evenals dat bij de gevolgen voor de patiënt het geval was onderscheiden we een medisch en een economische benadering van de *behoefte aan een technologie*. In beide gevallen wordt behoefte naar het oordeel van de patiënt zelf als een onvoldoende valide hulpmiddel voor MTA beschouwd. De kwantitatief geschatte verwachte winst in gezondheidstoestand, bepaald op een objectieve wijze vormt de basis voor beide benaderingen van de behoefte.

De *analyse van onzekerheden* maakt onderscheid tussen huidige en toekomstige onzekerheden. Eerstgenoemde onzekerheden blijken van verschillende aard te zijn: zij bestaan uit onzekerheden door onvoldoende informatie, uit 'intangibles' (moeilijk te meten zaken) en uit 'imaginables' (zaken die vanuit conceptueel oogpunt moeilijk zijn) . Toekomstige ontwikkelingen zijn per definitie onzeker, maar meestal zijn ze deels met enige zekerheid voorspelbaar. Dit wordt wel aangeduid met autonome trend. In aanvulling hierop kan analyse van diverse policy's plaatsvinden om inzicht te geven in de gevolgen van veranderingen in de omgeving waarin de technologie wordt toegepast. De kwantitatieve benadering van deze verschillende vormen van onzekerheid is goeddeels gelijk en maakt gebruik van kwantitative modellering.

Aan twee technische aspecten van de voorgestelde kwantitatief-empirische benadering wordt aandacht besteed: formele synthese en subgroep analyse. *Aggregatie-technieken* in het kader van MTA zijn bekritiseerd, meer in het bijzonder het toepassen van multidimensionele gezond-

heidstoetand-maten, het rekenen met individuele scores uitgedrukt in deze en andere gezondheidstoestand-maten en het berekenen van samenvattende maten zoals een kosten-effectiviteitsratio. Alleen al op methodologisch gronden is aggregatie als routine data-analysetechniek gerechtvaardigd. Daarnaast is zij een politiek desideratum.
De politieke aspecten van *subgroep analyse*, een tweede routine data-analysetechniek in empirisch onderzoek, worden aangegeven.

De hoofdstukken 4 tot en met 6 gaan vervolgens dieper in op de achtergronden van overlevingsanalyse, analyse van de gezondheidstoestand en de bepaling van de behoefte. Om overlap met de appendix te vermijden ligt het accent op methodologische achtergronden, aangevuld met een korte beschouwing over de empirische ervaring zoals deze o.a. in de appendix wordt gepresenteerd.

Hoofdstuk 4 beschrijft *overlevingsanalyse* in MTA. Beargumenteerd wordt dat deze in hoofdzaak bestaat uit sterk doorgevoerde toepassing van reeds bestaande technieken. Aanvullende berekeningen zijn noodzakelijk om gewonnen (of verloren) levensjaren als economische uitkomstmaat te verkrijgen. De uitdaging in menige MTA wat betreft overlevingsanalyse is gelegen in het bepalen van de overlevingsgevolgen van een interventie als geen informatie beschikbaar is afkomstig uit gerandomiseerd gecontroleerd onderzoek; deze situatie komt bij nieuwe technologieën dikwijls voor. Diverse benaderingen van de controle-groep overleving ('schaduwoverleving') in deze situatie worden beschreven, waaronder de toepassing van prognostisch modellen. De technische kenmerken van deze laatste methode, die kwantitatieve voorspellingen mogelijk maakt, worden beschreven. De resulterende voorspellingen kunnen zowel op individueel als groepsniveau worden gebruikt. Ze maken, mits op juiste wijze wordt rekening wordt gehouden met zgn. gecensureerde

waarnemingen, ook de berekening van de gewonnen levensjaren als gevolg van de nieuwe technologie mogelijk.
De ervaring in de levertransplantatie studie heeft duidelijk gemaakt dat schatting van de schaduwoverleving inderdaad mogelijk is als:
a. quasi-experimentele controle-groep gegevens beschikbaar zijn, of
b. interventie uitstel (b.v. wachtlijst) gegevens beschikbaar zijn, of
c. historische controle-groep gegevens beschikbaar zijn.
Aanvullende prognostische modellering is meestal mogelijk indien de bij a., b. en c. genoemde gegevens op individueel niveau beschikbaar zijn. In alle gevallen is homogeniteit van referentie en index patiënten met betrekking tot diagnosis en behandelingsvooruitzicht van belang.

Hoofdstuk 5 geeft enkele methodologische beschouwingen over het meten van de gezondheidstoestand, zowel vanuit een descriptief en als een waardetoekennend (in economische zin) gezichtspunt.
Allereerst wordt aan ziekte-aspecifiek gezondheidstoestand aandacht besteed. Betoogd wordt dat gezondheidstoestand een niet direct waarneembaar, theoretisch concept is, vergelijkbaar met tal van klinische concepten. Het bepalen van de validiteit vereist in dit geval a. consensus over een theoretische structuur die het betreffende concept bevat en b. consensus over te voorspellen feiten voortvloeiend uit dit concept dan wel voortvloeiend uit gespecificeerde parallelle of anderszins gerelateerde concepten.
Voor de *beschrijving* van de gezondheidstoestand wordt een algemene structuur voorgesteld. Indien vervolgens hierop geënte vragenlijsten worden ontwikkeld, dan kunnen zowel het statistische efficiëntie criteria als inhoudelijke criteria hierbij behulpzaam zijn, b.v. bij de keuze van dimensies, items en antwoordmodaliteiten. Beide criteria hebben hun voor-en nadelen. Vervolgens is een aantal criteria beschikbaar om be-

schrijvende instrumenten op bruikbaarheid en kwaliteit te beoordelen. Binnen de context van gezondheidstoestandbeschrijving in MTA heeft, zo wordt aangegeven, 'tijd' een belangrijke conceptuele positie, zich uitend in vraagformulering en interviewfrequentie.
Onze ervaring met de descriptie van de gezondheidstoestand in de Nederlandse transplantatie MTA's wordt kort beschreven. Twee kenmerken van deze descriptieve analyse lijken binnen MTA van meer algemene aard te zijn: de moeilijke statistische analyse van links en rechts gecensureerde gegevens en - als bij overlevingsanalyse - de problematische vergelijkende analyse als een experimentele opzet afwezig is.

Het tweede deel van het hoofdstuk wordt gewijd aan de meting van de gezondheidstoestand als onderdeel van de *economische meting van de patiënt gevolgen*. Tot nu to is de beste beschikbare techniek om tot een economische waardering van de patiënt gevolgen te komen de berekening van gewonnen levensjaren, of, indien mogelijk, van kwaliteitsgecorrigeerde levensjaren (ook wel aangeduid met het acroniem 'QALY's' voor Quality Adjusted Life Years's). Ten behoeve van de berekening van QALY's zijn a. meting van de gezondheidstoestand van patiënten, b. destillatie van gezondheidstoestandbeschrijvingen en vervolgens c. waardering van deze gezondheidstoestandbeschrijvingen drie noodzakelijke stappen. De toepassing van QALY's in zogeheten kosten-utiliteits-analyses berust op vele assumpties, die zorgvuldig ontworpen meetinstrumenten vereisen. Wat dit laatste betreft zijn wij erin geslaagd om met buitenlandse collega's een instrument voor gezondheidstoestandwaardering te ontwerpen (EuroQol), dat internationaal beschikbaar is. Met dit instrument konden wij de geldigheid van een essentiele aanname voor QALY-berekeningen bevestigen: de interindividuele consensus met betrekking tot de rangordening naar wenselijkheid van een aantal complexe gezondheidssituaties van overigens gelijke duur.

De resultaten van de kosten-utiliteits-analyse in de Nederlandse transplantatie MTA's ondersteunen de haalbaarheid van de QALY-procedure.
Als een apart onderwerp gaat Hoofdstuk 5 in op de veelbelovende ervaring met het personal computer ondersteund enqueteren van patiënten.

Hoofdstuk 6 besteedt aandacht aan de *bepaling van de behoefte aan een specifiek technologie*. Conceptuele keuzen en dataschaarste kenmerken dit onderwerp. Allereerst dient tussen een definitie gebaseerd op verwachte gezondheidsverbetering of één van de definities gebaseerd op andere principes worden gekozen. Deze keuze is dezelfde als die aangeduid in Hoofdstuk 3 bij de definitie van rechtvaardige verdeling ('equity'). Naar ons inzicht dient het begrip behoefte uit te gaan van de eerste definitie. Vervolgens beschrijft het hoofdstuk een medische en een economische uitwerking van de eerstgenoemde behoefte definitie. Een empirische behoeftebepaling kan vruchtbaar gebruik maken van het onderscheid tussen 'incidentie-technologieën' en 'prevalentie-technologieën'. In vervolg hierop worden suggesties gegeven voor de bepaling van de behoefte van de vier onderscheiden technologische hoofdtypen.
Verslag wordt gedaan van de schatting van de behoefte aan levertransplantatie. Deze schatting volgens eerdergenoemde principes was dankzij enkele gunstige omstandigheden goed mogelijk.

Hoofdstuk 7 keert terug tot de drie eerdergenoemde kernvragen.

Wat is de wetenschappelijke basis van MTA?

Het zal duidelijk zijn dat in dit proefschrift de stelling verdedigt dat er een wetenschappelijke basis bestaat voor empirisch MTA-onderzoek.

Rigoreuze standaardisatie, kwantitatieve modellering en integratie van medische en economische concepten zijn de ingrediënten voor de methoden van MTA. Voor de verschillende onderwerpen zijn standaardtechnieken en -hulpmiddelen beschikbaar, of, minstens zo belangrijk, standaardmethoden om deze te creëren. De methodologische keuzeproblemen, die soms onmiskenbaar aanwezig zijn, zijn zowel het gevolg van een levendig multidisciplinair wetenschappelijk debat als van de samenhang tussen wetenschapsbeoefening en maatschappelijke besluitvorming.

Welke omstandigheden begunstigen het gebruik van MTA?

Het gebruik van empirische MTA-resultaten hangt van zowel politieke als wetenschappelijke randvoorwaarden af. Met betrekking tot de politieke randvoorwaarden is het duidelijk geworden dat een rationeel en expliciet besluitvormingsproces bijdraagt tot vruchtbaar gebruik van MTA-informatie. In de Nederlandse situatie heeft de Ziekenfondsraad een besluitvormingsprocedure voorgesteld die sterk gelijkt op die is beschreven in Hoofdstuk 2. Hoewel ze nog niet volledig geëffectueerd is, heeft de voorgestelde werkwijze in combinatie met de drie proef-MTA's (over in vitro fertilisatie, hart- en levertransplantatie) al positief bijgedragen tot de besluitvorming over deze en andere nieuwe technologieën.
Daarnaast zijn de voorwaarden voor MTA-onderzoek van belang, zoals het bestaan van onafhankelijke multidisciplinaire MTA-instituten, de politieke steun voor deze instituten waar het gaat om het toepassen van standaard methodologie, en een goede aansluiting van klinische ontwikkelingsprogramma's aan het MTA-informatie probleem.

Hoe kan MTA in de toekomst verder worden ontwikkeld?

Het antwoord op de laatste vraag over de *toekomstige ontwikkeling van MTA* heeft duidelijk een politiek karakter. De eerdergenoemde voorwaarden zullen ongetwijfeld haar ontwikkeling beinvloeden, maar ze zijn in onze ogen niet allesbepalend. Belangrijker achten wij de twee volgende, algemene ontwikkelingen.

In de eerste plaats de ontwikkeling van een wetenschappelijke relatie tussen MTA-instituten en haar methodologische en klinische buren. In de tweede plaats de ontwikkeling van een politiek momentum dat erkent dat MTA een continue activiteit dient te zijn, als antwoord op de in economische zin asymmetrische relaties in de gezondheidszorg. Wij verwachten dat deze laatste ontwikkeling uiteindelijk afhangt van de wijze waarop de emancipatie van de medische consument vorm krijgt.

NOTES

Chapter 1

1. This description describes changes in hospital function in this century only (Relman, 1988; Foucault, 1975 [p. 54]).

2. An example of moving borders of professional territoria may be found in gynaecological surgery. At the end of the 19th century gynaecologists did discuss the desirability of including surgery of the female genital tract into their disciplin. The answer is known, more than the arguments. Currently the discussion is renewed in the question about the prefered surgeon in surgery of the female urogenital tract. Technological advances (diagnostics, therapeutics) and professional specialization (urology, born from general surgery) gave rise to the current debate (De Kroon, 1990).

3. In this thesis the third world countries are excluded from analysis only for practical reasons. Their importance is acknowledged (Banta, 1984).

4. We are aware of other factors inducing or influencing changes in health care. The relative impact of technologically induced changes may be quite moderate as is illustrated in our study on Future Scenarios of Obstetrics, Gynaecology and Reproductive Medicine (De Kroon, 1990).

5. Registration may be intentional or casual (e.g. the result from national price lists for reimbursement or budget arrangements). It may be the result of a statutory requirement but also the result of initiatives of companies or consumers
   A brief overview of the Dutch state of affairs may serve as an example. In the Netherlands a complete drug registration is counterbalanced by the absence of an analogous registration of devices and medical equipment (only coming to public attention when a particular type of artificial heart valves proved to be unsatisfactory). Health insurance registrations and and hospital administration systems usually respond with some delay to new medical procedures. Organisational changes within statutory limits are short of registration - as may be illustrated by the "hidden" existence of transplantation teams (see also Geleijns, 1990).
   Though registration usually will be pluralistic, even within one country, and will apply to particular technologies (e.g. drugs), we conclude that scarcety of resources is but one of the factors which may force the medical system to adress the problems of emerging technologies.

6. In this alinea, the emphasis is on medical technologies influencing society. Examples of the reverse relation are more difficult to determine, but may be found when looking for societally demanded technologies. On a more abstract level Habermas (notably in the essay "Technik und Wissenschaft als Ideologie"), and to a lesser extent Illich, Foucault and de Swaan have elaborated on this subject (Habermas, 1968/1976; Illich, 1974/1975; De Swaan, 1984).

7. See Table 1.1 for circumstancial factors contributing to recognition.

8. We do not believe - from a historical point of view - this response can precede technological change, which is a basic assumption underlying the philosophy of societally guided technological change. Organizational arguments could be added (Mintzberg, 1979 [particularly p.371-379]).

9. As will elaborated on later, a new technology is defined as 'new' from the perspective of the particular organization involved; 'technology' refers to a health care programme - in our case for patients with end-stage liver disease from non-alcoholic origin.

10. As in farmacology, 'side effects' as opposed to 'main effects' have to do with the intention, not with the comparative impact of the effects.

11. 'The technological imperative' is sometimes used as a - in our view - theoretically unfruitful concept to describe the autonomous character of occurence and diffusion of new technologies. See Relman (1988).

12. A similar, analogous change may be observed in the debate on environmental pollution, though the roots of this debate are older (see the history of the American Office of Technology Assessment in e.g. Banta, 1990).

13. The emphasis on the decision process on the the macro level in this section implies it to be particularly relevant for the first group of parties of Table 1.2. However, with adaptions according to the specific interests, the body of this section also applies to most of the other parties (IOM, 1985 [p. 26-]). Separate attention will be paid on this adaptations, particular in the description of the decision cycle.

14. At this place we will not elaborate on other responses of health care authorities on new technologies.

15. The development of such a rational and intelligible approach of complex problems should not only be described as the concerted action of responsible actors in the field, but also as a result of the gradual transformation of the former politization of (medical) science into scientification of (health) policies - see also e.g. Habermas (1968/1974).

16. This relation is usually neglected when parts of the structure, process or outcome of the decision process are held up as an example to another country.

17. About two thirds of the population is covered by this scheme.

18. In *orthotopic* liver transplantation the diseased liver of the recipient is removed (explantation) and subsequently replaced (implantation) by a liver of approximately the same size from a donor with the same blood group. Few advocate *heterotopic* liver transplantation in some cases. This technique differs from the previously described technique in the apposition of the donor liver, rather than the replacement of the diseased liver.

Chapter 2

1. Despite a uniform agenda of topics to be judged, there is a wide array of decision processes in Western countries to be observed. Usually the nature of the technology and not that of the health care system is decisive for the nature of the process (e.g. in most countries new drugs are under the same vigorous control). See also OTA (1980), Banta (1982) and Jennett (1986).

2. In the Netherlands the National Health Insurance Board (NHIB; Dutch: Ziekenfondsraad) has deviced a similar decision flowchart, combining political and informational elements. From 1983-1985 the author as a member of a working committee of the NHIB was closely involved in the development of this flowchart and its subsequent implementation.
Though the resulting cybernetic cycle presented here is in some aspects deviant from the flowchart of the NHIB, this section owes much to the participation of the author in the discussions of the former working committee of the NHIB on this subject. See a.o. NHIB (1983, 1986).

3. A good example in the Netherlands of the lack of impact of early warning on the application of a new technology was pointed out by Rigter (1989). The wealth of committees and working parties on genetic testing, which exist since 1984, sofar has not influenced much the introduction and early adoption of this technology (see Rigter 1989 [p.40]).
Moreover we think, supported by history, that the primary impetus of a decision process on the application of a medical technology is the *actual* effect of a technology, not its *potential* impact.

4. Maybe these early warning systems have some limited use for the decision processes on budgets for research and development. In that case the decision-cycle should be adapted accordingly.

5. We are aware the judgments on relevance and information surplus require at least some information to be available.

6. We distinguish the following criteria:
   1. maximal cumulative cost per person;
   2. maximal cumulative cost per year of the programme;
   3. minimum cumulative effectivity per person;
   4. minimum cumulative effectivity of the programme;
   5. maximal marginal costs compared to minimal marginal effects (comparison by division) per person;
   6. average cost-effectiveness of the health care programme;
   7. direct or indirect consequences with influence on future generations;
   8. direct or indirect consequences which imply interpersonal weighting of different interests.

   These eight criteria may serve as a checklist.
   Criteria 1. and 2. are input-criteria.
   The first criterion is related to the end-to-means issue. It compares costs of the target technology to the doing-nothing alternative. The second is related to the issue of a fixed budget for health care.

The first criterion may be simply expressed as total direct costs per person or, maybe better, total costs per person. A critical calue will be difficult to find from history, taking the acceptance of heart transplantation and the existance of units for permanent ventilation of patients needing life-long respiratory assistance into account. Critical values outside the health care system do exist (Jones-Lee, 1976; Mooney, 1977; McGuire, 1988 [p.87, 88]).
The second criterion may be best expressed as a percentage of health care costs or as a percentage of costs within the associated speciality. Note that a criterion of total specialty costs of e.g. 1% will not easily be surpassed.

Criteria 3. and 4. are output-criteria.
The third criterion is related to what some have mentioned 'medical success'. Obviously from a logical position we have to compare the effects with the doing-nothing alternative as we did in criterion 1. with costs. Quantification of the natural history will frequently be difficult (See Chapter 4). Again this might be regarded as an end-to-means issue. 'Alternatieve geneeskunde' will satisfy criterion 3. regardless the threshold chosen.
The fourth criterion relates the importance of an eventually new programme to the cumulative benefit of the specialty involved or even health care as a whole.
Elaboration of the second and fourth criterion needs a national cost-of-illness matrix together with a national yield-of-health care matrix (Koopmanschap, 1991; Health Council, 1986 [p.122, 123]).

Criterion 5. and 6. are aggregated criteria i.e. they combine information on different characteristics. The fifth and the sixth criterion are about end-to-means, relating information on ends (criterion 3) and means (criterion 1) and - this is an extension of the concept end-to-means in its usual interpretation - relating it to similar information on the best alternative action. Thus they are also economical criteria, particularly the sixth.

The last two criteria are ethical. We are aware one may reject the existence of ethical criteria which are derived from ethical principles (the ethical products) and may prefer the ethical role to be represented by critical reflection on the acceptability of new technologies (the ethical process) (Ten Have, 1988; see also Health Council, 1986, Anonymus, 1988).
The seventh criterion is about technologies which directly or indirectly produce genetic information and about technologies in health care programmes with effects on the population at large (e.g. disease eradication programmes).
The eighth criterion is relevant if technologies implicity or explicity violate personal interests of non-diseased. Quarantaine is an example which gained renewed interest during the AIDS-epidemic. Vaccination programmes with their inherent victims of rare complications are but another. Screening on cancer of various kinds should be mentioned but for the same logic the shift from hospital to home care.

7. Dutch health care authorities speak of five questions, which are of relevance in the context of this decision process on the uptake of new technologies in health care. Adapted to a technology decision (Anonymus, 1988; Health Council (Gezondheidsraad), 1986). These questions are:
   a. Can the technology be regarded as health care?
   b. Is the technology ethically acceptable?
   c. Is the technology effective, is the technology efficient?
   d. Can the technology be paid for?
   e. Does the technology belong to the realm of the constitutional duties of the state?
   Operationalization of these questions has been started on. Answers sofar are lacking, which makes us aware of a fundamental void in the decision chain.

8. These agendas may be e.g. be derived from the issues of the Journal of Technology Assessment in Health Care. We are aware of the different agendas in developing and industrialized countries (Banta, 1984).

9. In the Netherlands the flowchart of the NHIB (Ziekenfondsraad) also starts with a recognition phase. The flowchart assigns a prominent role to both health care providers and consumers, followed by most of the parties mentioned in Table 1.2. See NHIB (1986, [p.15, 23]).

10. An interesting example exists in the Netherlands which shows the difficulty of guaranteeing this type of access. The Steering Committee on Future Scenarios in Health Care (Dutch: Stuurgroep Toekomstscenario's Gezondsheidszorg) which started in 1983 and is closely related to the Department of Health, acts both as a generator and a mediator for investigational projects on future health (care) issues (STG, 1987). A particular generating role for persons and parties outside this committee is explicitly stated. Only recently a project has been undertaken, which indeed was initiated from outside (Gynaecology, Obstetrics and Reproductive Medicine, see De Kroon, 1990).

11. Within the Dutch system the Health Council and the NHIB (Ziekenfondsraad), given the preceding criteria and given their constitutional position in the health care system, are the natural agencies to select technologies to be judged (Health Council, 1986 [p.81-84]).

12. In this context the NHIB (Ziekenfondsraad) flowchart mentions two criteria.
The first is *maturity* ("a minimum level of consensus on medical and technical aspects of the technology involved"). The second regards *acceptability* ("no unsurpassable objections from a societal, political, ethical or legal point of view").
Surprisingly a criterion of *relevance* is lacking. This omission avoids conceptional difficulties (what is a 'new' technology) and political difficulties (what is an 'important' impending technology) in the first phase of the NHIB flowchart. We think this omission is justified in this case due to the existence of a variety of well-functioning sounding boards. See NHIB (1986 [p.16]).

13. In the NHIB (Ziekenfondsraad) flowchart the framing of the decision problem is slightly different from the definition that described here (NHIB, 1986 [p.16-]).
In the NHIB a potential shortcut in the flowchart is introduced, after maturity and acceptability of a new technology have proved to exceed minimum standards. When costs of the new technology seem to be lower than previous technologies and *all other things* are equal, a simple decision is possible, resulting in full support to the new technology. If not, comprehensive judgement is necessary. Thus the decision process of the NHIB consists of two consecutive decision 'nodes', the first devoted to cost balancing alone, the second to weighing of all aspects. In a technical sense this implies Phase 3 to 5 to be executed twice, if simple substitution is not to be expected.
We doubt the argument for such a shortcut to be valid, as the expectation of pure substitution is not justified both on theoretical and empirical grounds:
   a. lower costs imply extended use; inevitably selection criteria will change (based on the fundamental assumption that the health care market is not satisfied and will always be subject to economical laws of scarcity);
   b. the utility of ignorance and the disutility of change of medical practice will usually keep alive existing technologies: full replacement is an uncommon event, partial replacement a more frequent one and parallel or serial application (additional use)

the rule; the last rule is especially valid when substitution is an interprofessional affair (see e.g. CABG versus PTCA in coronary artery disease);
c. *all other things* are only rarely equal and equality is not easily verified.

14. Frequently the bringing-under-control (an objective related to the structure and process of the technology) has priority over the control itself (an objective related to the outcome).
We will give an example of this change from a public health objective towards management objective. Most of the parties, which were involved in the liver and heart transplantation technology assessments in the Netherlands, perceived the studies as instruments to bring an otherwise autonomous development under public control. The freezing of these programmes cannot easily be explained as a necessary condition to be met in order to carry out a technology assessment.

15. For whom refers to indication and selection. By whom refers to profession, setting and may be extended to particular persons and/or institutions.

16. 'Criteria' refer to some critical level, either separately or combined, of these issues. A critical level implies measurability. In its simplest form it is absence or presence. Most issues should be described more subtle.
The use of multiple criteria is discussed in Chapter 3. Several formal and informal techniques exist to reach a conclusion based on aggregation of heterogeneous issues. The choice of the aggregation technique is primarily a political one and should ideally reflect major societal values. Aggregational techniques differ with regard to principle (outcome or process oriented - teleological versus deontological), consistency (if non consistent, the same principle may yield different decisions in the same situation) and verifiability (closely related to explicity and measurability).

17. A description of intangibles is not enough in that case. The development of criteria is the core of the problem. To declare these type of issues 'ethical' does not solve the problem if explicit valution is aimed at.

18. We e.g. did not succeed to get access to the data of the European Registry of Liver Transplantation.

19. We can define the effectivity space and the cost space as the estimated difference of effects and costs between the introduction of the new technology vs. the null scenario for, say, 5 years; with old technologies we can use alternative measures of potential cost impact (Phelps, 1990).
In liver transplantation the MTA-costs accounted for less than 5% of 5-year cost space (see *A10*).

20. The NHIB (Ziekenfondsraad) flowchart gives little information on the contents of this part of the process.

21. The NHIB (Ziekenfondsraad) developed a system of provisional financing of new technologies, provided the particular health care providers carried out investigations to support a decision of the NHIB in the context of the flowchart ('Ontwikkelings-geneeskunde'). From the absence of Phase 2 and particularly of Phase 3 in this system we conclude the following:
    a. health care authorities are reluctant to take up their political role particularly in Phase 2; issues and aspects are not clearly specified and the relevance of projects for the health care system at the national level is uncertain;
    b. the managing objective appears to dominate the judging objective; the information to be generated in these projects is not clearly defined; this might result in a a misfit of need and supply of information at the end of these projects; from a managing perspective this misfit is of less relevance.

22. If medical technology assessment serves other purposes as well (see footnote 21 about the bringing-under-control objective of health policy makers), this statement should be refined. In that case decisions are made during the investigative process of Phase 4, and the role of health policy makers is extended accordingly.

23. The NHIB (Ziekenfondsraad) flowchart clearly fits in this scheme, though two decision nodes are involved, the first restricted to cost balancing (note that the cost definition is unknown). See NHIB (1983, 1986).

24. Undesirable according to criteria not specified ahead.

25. In the NHIB (Ziekenfondsraad) flowchart Phase 6 is lacking, which is surprising in view of the intention to include already existent technologies in the flowchart. Some examples of monitoring do exist. Two will be mentioned.
    The Dutch Hospital Provisions Act includes a section (Art. 18), which enables application of specific medical technologies to be restricted and - since 1989 - to be monitored.
    In a related field the Dutch Steering Committee on Future Health Care Scenarios usually institutionalizes monitoring following a research project on health care developments. These projects do not investigate particular technologies but intend to forecast on a time horizon of about 15 years developments in broad areas like cardiovascular diseases, cancer and accidents (Van Beeck, 1989).

26. Other reasons may be e.g. extreme scarcity, leading to implicit or explicit selection processes, and geographical variations of application, leading to unequal access for the patient.

Chapter 3

1. Social issues seem the victim of semantic difficulties. Sofar they have turned out to be either a. issues dealt with in sections 3.3 to 3.7, b. ethical issues or c. issues outside the realm of MTA related to other phases of the decision cycle or to the decision cycle as it stands.

   Social impact of technologies is more conveniently dealt with in sections 3.3 to 3.7. The topics 'impact on process of health care delivery at large', 'impact on public health', and 'impact on health care expenditures and GNP', etc. have a natural place in these sections.

   Sometimes problems are called 'social' when societal values are at issue. In that case we would prefer to call these 'economic' or 'ethical', as appropriate.

   Finally social issues may refer to other phases of the decision cycle e.g. to the social process of decision and implementation in Phase 5 of the decision-cycle. Though important we regard this elaboration a corpus alienum in the information phase. The role of MTA in the preceeding or subsequent phases, the role of other factors influencing the decision cycle, or the rationality or desirability of such a technocratic system in it own all have their own meta-analytical place.

2. Randomization in experimental designs pertains to two aspects in the experiment. First a random *sample* is drawn from the target population. Later, the *intervention(s)* are randomly allocated. Obviously the purpose of randomization is to enable generalization and unbiased comparison respectively.
   The recruitment and selection of patient's in clinical trials usually do not satisfy the first requirement, limiting the scope of the analysis to statements about within-trial differences (Senn, 1990). Apparently this requirement is of limited perceived interest.
   The second stage is generally considered to be a necessary attribute for any RCCT worthy of the name. However, pseudorandomization in RCCT's of new technologies is not an uncommon event. In that case one or more of the essential requirements of this procedure are compromised. E.g. randomization occurs prior to the doctor's request for participating in the trial and the allotted treatment alone is offered to the patient with an adapted informed consent. The scientific and ethical problems of this adaptation are self-evident.
   One variant of the treatment allocation randomization should be mentioned. A first randomization towards a RCCT or a preferred treatment protocol followed, in the case of a RCCT, by random allocation of one of the interventions has been advocated.

3. Sometimes it may be difficult to device a control arm which matches all requirements: e.g. complex mimicri-devices to assure blinding are unsuitable for the determination of control cost.

4. Only two arguments againt RCCT's seem justifiable.
   First of all practical difficulties may be too difficult to overcome, though it is the blinding and not the randomization which frequently is difficult to achieve. Problems may arise when one out of several best alternatives can only established after some trial and error due to variability of response (e.g. anti-depressive drugs, heart failure drugs).
   The other difficulty arises from ethical considerations. The dominant ethical position leaves little room for experiments when significant benefit may expected from the new technology. As few clinicians would be motivated to undertake experiments if this expectation was absent, satisfaction of the ethical imperative to give the best to the individual, would result in a moratorium on RCCT's (see e.g. Mosteller, 1981). The resulting return to the pre-Baconian era of the alternative healers in our view is unacceptable.

5. The interests of national health insurances only partially parallel those of private health insurance companies; the interests of specialist departments do not coincide with that of the hospital at large; and the interests of individual patients and categorical patient organisations are not identical to the interests of consumer organisations.
   The case of the three Dutch MTA's described in this study is interesting as a mix of interests may be observed among those supporting these MTA's.

6. A specific trend is clearly visible in the MTA activities of OHTA. The more recent the MTA, the more specific are the patients described (IOM, 1985 [p.357-359).

7. Another reason is the frequent relevance of including organisational aspects in the description. Lack of standard methodology in this respect may be observed, e.g. on the issue of centralization versus decentralization, the issue of decision-making in complex treatments, etc.

8. A health care programme is defined as the combined professional effort to comfort the need for medical services for a particular patient group (see also chapter 2).
   A patient group is defined, like the DRG-methodology, as a group of patients with a similar pathologic condition and with a rather homogeneous pattern of medical consumption.

9. One exception is the appearance of a new technology being considerable cheaper than existing alternatives. Examples are abundant in the field of pharmaceutical products. In most cases it is the pricing, not the profile, which is the interesting feature of a new drug (me-too preparations and, specifically in the Netherlands, FNA-preparations to replace brand name drugs). We are aware of the commercial aspects of undertaking CEA's and even full-size MTA's on these occasions.

10. We disregard an 'essentialistic' approach.

11. In clinical studies, outcome usually is defined by disease-specific (process) parameters which are derived from the pathological concept of the disease (e.g. ejection fraction and cardiac output in cardiac insufficiency, FEV_1 in lung insufficiency, blood chemistry in liver insufficiency and renal insufficiency, etc.). Parameters reflecting the degree to which signs and symptoms exist may be added, when they are thought to sufficiently valid.

12. As in the exact sciences, these type of parameters may be only defined by means of their measurement procedure.

13. Particularly if clinicians use a carefully designed set of disease-specific parameters they put questions on the value of these disease-non-specific outcome parameters, which sometimes are quite insensitive from their point of view.
Obviously these questions are only raised if the choice *can* be made i.e. if a disease-specific technology is looked at. In fact most technologies are applied to patients with different disease-categories. In those cases the general level *must* be chosen, as a disease non-specific decision is usually aimed at. Hence we regard disease-specific outcome-measurement a special case within the medical concept of outcome analysis.
The emphasis on disease-non-specific outcome measures easily gives rise to conflicts about the design, the operationalization and the publication of the MTA-study as they compete with the clinical investigation of the programme from which data are drawn.

14. We disregard other unidimensional parameters for their limited relevance, both from a theoretical and practical viewpoint. See also Drummond (1987 [p.15]).
The broader concept of outcome, which is adopted in QALY's and HYE's (see Chapter 5) may include pain or disease avoided but also reassurance gained by information or emotional support and other process utilities; McGuire, 1988 [p.46-51,92,93,106]).

15. For analogue reasons consequences have to be valued separately, see section 3.4.

16. For a brief account of the technical procedure see Drummond (1987 [p.48-53]).

17. The intensity of patient-monitoring which usually is defined from a research perspective should be adjusted to reflect intensity under normal conditions. Treating the costs of any *fixed* follow-up routine - not depending on patients complaints, etc. - as research costs yields an upper bound of treatment costs.

18. The paradox arises if we aggregate on a cross-sectional base instead of a longitudinal base.

19. In the typical case the degree of uncertainty determines the decision to be made, regardless the actual values of the estimated range of plausible results.

20. In many cases a clear age-dependency exist which justifies demographical trend analysis. If constraints are likely - the relevance may be small.

21. Here we disagree with Hoogeveen (1989 [p. 122]).

22. Marginal analysis is an analytical technique which examines the relationship between incremental changes in investments in a product (input) and incremental changes in output (Luce, 1990 [p.26-28]; examples in Van Oortmarssen, 1990 and Neuhauser, 1975).

23. The first step in the NHIB (Ziekenfondsraad) flowchart in (see notes in Chapter 2) implies some form of threshold-analysis.

24. In practice deterministic models usually are easier in providing thresholds than probabilistic (syn. stochastic) models, but feasibility has its price.

25. Note that this is a socalled stochastic model.

26. Within the NHIB we remember an analysis of concordance been tried out.

27. Counting over time however does not necessarily mean one event has the same significance for the individual at two occasions. The second anginal attaque may be born easier or just the reverse. But for their prevention we treat them equal.
Counting over different effects does not mean the effects are identical. The only share the characteristic that their relative value may be known. If an apple does $ 0.25 and a pear $ 0.50 it does not mean two apples *are* a pear.
Counting over individuals does not mean they are identical but assumes for the aspect chosen, abstraction is allowed.

28. Available data lend themselves not very well for comparison in most of the cases, due to departures from the basic rules of cost-effectiveness analysis (Rutten, 1989). Even if the analysis has been carried conform these rules it should be particularly clear whether *marginal* or rather *average* cost-effectiveness ratios are compared. This issue is closely related to the choice of the alternative action in the analysis.

29. For example in the Netherlands these implicit criteria have decided on the institution of suboptimal policies in cervical cancer screening and the temporary exclusion of liver transplantation in children from the insurance package despite excellent performance as defined by medical and economical figures.

30. Preferably in the third phase of the decision-cycle (see chapter 2).

31. In analyzing the therapeutic effectivity of liver transplantation (parameter: survival) the variable 'clinical center' appears an important predictor variable only second to 'pre-transplantation severity of disease'. As this outcome might result in licensing only clinical centers with best performance, the health care providers do not agree with the analysis.

32. At this place we will not elaborate on study-specific details. Generally the arguments may be analytically described as:
    a. an explanatory variable violates professional practice interests;
    b. an explanatory variable violates scientific interests;
    c. an explanatory variable violates financial interests;
    d. an explanatory variable violates societal interests, particularly the pursuit of equity.

33. Thus we speak about a consequentialistic approach.

34. Except for the Rawlsian definition (Rawls, 1972). The nice theoretical operationalisation of this problem of Blackorby (1977) should be mentioned. However, data requirements are immense.

35. Usually a linear function combining X and Y is chosen. We prefer the general form Y = f(X). This enables for example discontinuous or non-monotonic functions.

36. Present health status as a instantaneous judgment on mobility, psychological status, etc. appears not to to be a valid measure: for example a disease like AIDS exerts little impact on actual health status, but few would support the view that this observation implies patients are in little need for care. Prognosis c.q. survival should therefore be included. This introduces a QALY-like estimate 'as an aggregate measure of expected misery' (De Jong, 1983).

37. The same holds for other operationalisation of need (see chapter 6).

38. Besides, the quantitative calculation of this relation will be difficult and results may be surprising. E.g. in liver transplantation, we found age as a strong negative predictor of survival without transplantation, and (within a rather broad range) as an irrelevant predictor with regard to survival with transplantation and quality of life before and after transplantation. Assuming *all* other things equal the efficiency criterion in this case would even favour the aged as would the vertical equity criterion, while horizontal equity considerations would show no preference.

39. Within economic evaluation the procedure inflates costs at an extremely high rate.

40. This includes the anencephalic discussion (Harrison, 1986).

41. Upon the request of health care authorities these issues were dealt with by expert committees of the Health Council. The use by health care authorities of information on legal issues from the Dutch transplantation MTA's (particularly from the Health Council, 1988; id. 1989) has been rather disappointing. The donor card issue still awaits elaboration and fruitful legislation. The more general problem of inevitable selection still has not been adressed in a satisfactory way.

42. *Phase 1*
Legal considerations are of paramount importance in the selection of technologies to be subjected to explicit decision-making.
Depending on current legislation on the approval of new technologies (Health Council, 1987 [p.104]) MTA-like research may be indicated.
Sofar legal considerations for the selection of technologies to be subjected to MTA have been more fundamental (ethical). Attention has been drawn particularly when technologies were expected to have impact on either the start or the end of life. This interest has some negative effects.
    a. Technologies, which may be defined as important based on other criteria may be neglected. Technologies associated with legal or ethical controversies may inadvertently be juridified: the euthanasia-question does not need a MTA of opiates.
    b. Even if a MTA is needed, the existing variety of fundamental legal positions may be projected into the information need, leading to incompatible questions and criteria to be used.

    Some have advocated a global judgment of the legal acceptability of a new technology, setting intended ends to means. History provides ample evidence for the relevance of this premature judgment, apart from the hazards of judgment of intentions. At this stage we think this to be the primary responsibility of professional

organizations. Of course this presumes mature professional organizations embedded in a consensual legal system as in most Western societies.

Phase 2
The definition of the decision problem is also related to legal notions but we do not agree with the proposed principle of current legislation and jurisprudence defining the decision problem. The reverse should be true: the decision problem should clearly state the potential relevance of outcomes to e.g. international regulations, and to legislation e.g. on the provision of health care. Particularly in new tecchnologies the time-lag of legal accomodation, though useful in another way, prevents the legal system from a prospective task of defining the decision problem.

Phase 3
The information need might or might not include full analysis of all legal consequences associated with any of the decisions to be made. We reject this broad view, preferring legal issues *in* a MTA to be restricted to those pertinent to the technology.
This preference is based on the following arguments:
a. The major part of legal consequences is about legal adaptations. Their analysis usually can wait. For this part the political decision is conclusive. Though the existence of fundamental legal issues should be noticed - see our examples on heart transplantation, liver transplantation, in vitro fertilization - these are usually part of the ethical issue.
b. An impressive list of additional externalities to a decision may be formulated (e.g. scientific spin-off). Full analysis of all externalities usually is unfeasible.

We think the legal view does ask for special emphasis on particular issues: description of the health care programme, survival analysis, analysis of need, subgroup analysis, equity considerations and legal issues, as defined before.

Phase 4
With regard to empirical MTA's we should mention the legal features of clinical experiments and the use of registry data. Clinical pilot programmes take with them 4 legal issues:
a. free choice of the patient to participate and eventually to leave the pilot programme;
b. limited access to the pilot programme
c. liability of the provider
d. confidentiality of patient's data.

Phase 5
The decision may call politicians, and legal, ethical and medical experts to arms.
The first appear to be the legitimate decision-makers.

Phase 6
Legal issues in implementation and monitoring are treated in section 2.7.

43. The proposition to allow for 'medical' selection criteria only is not a unambigious solution. This proposition simply moves back the problem of selection from the legal to the medical arena, suggesting the former is of little help. In addition it deliberately avoids a clear operationalization of the adjective 'medical'. At least three different meanings are conceivable.
If 'medical' means 'to be determined by a medical expert' we arrive at a *process* criterion, ascribing the legal power to select to doctors only. This process-definition naturally allows for non-unique explanations. Two examples: 1. the choice for a selection criterion is determined by a medical expert or 2. the actual level of a predefined criterion is determined by a doctor.
If 'medical' means 'related to the outcome of the disease', we arrive at an *outcome* criterion. Its operationalization asks for variables predicting the course of disease and the degree of succes of treatment. 'Medical' in this sense includes age as a justified variable as age usually reigns all aspects of the disease (incidence, progress, susceptibility for treatment). Judging from the context of the proposition (medical criteria as opposed to the ace criterion) we doubt this meaning of 'medical' has been meant.
Some have proposed a lottery instead of medical selection in case of inevitable selection. Though attractive, this usually is not defensible (from the viewpoint of equity and the viewpoint of efficiency) when life-saving technologies are involved. Those with the 'best' survival without the target technology - unrelated to the net benefit - are unduly favoured.
We conclude a legally acceptable formula has yet to be found.

44. "By the end of the 1970s there was a gradual change in ethical discussions. The importance of general philosophical issues for medicine was recognized, and consequently the need for a philosophical basis for health care ethics. Analysis of moral dilemma's inevitably raised questions about the nature of health and disease, and the goals of medicine. The moral problems of contemporary biomedicine were in this phase linked with a philosophically inspired taxonomy of topics, including autonomy, personhood, justice, coercion. Ethical evaluation of the doctor's thoughts and actions would be more penetrating when attended by an analytical dissection of his/her presuppositions and by a critical examination of conceptual foundations of clinical practice. (...) Once we have chosen a conception of medical ethics as a process, the question arises of how moral reasoning proceeds and what purpose it has in medical matters. (...) The method of medical ethics [as process GJB] has at least three steps: clarification, analysis and recommendation."

Chapter 4

1. The technical expression is 'terminal event'.

2. The present waiting-list experience of heart transplantation patient indeed seems to confirm the bold and, according to clinicians, unlikely predictions (in 1988) of about 25% 1-year survival without heart transplantation.

3. The phenomenon is well-known in psychological literature.

4. The pathogenic profile of organisms may change, but also the profile of the host may change.

5. This temporal drift is e.g. a serious threat to the valid use of the Framingham-data.

6. We do not advocate wating lists, to be clear.

7. From the Dutch heart transplantation programme and from the ET-Newsletter we know that these conditions are not always satisfied in organ transplantation waiting lists.

8. Note that this basal hazard function is not equivalent to the hazard function of the normal population. It contains observed homogeneity (e.g. the diagnosis of all patients; this cannot explain survival differences in *this* group), observed heterogeneity (measured factors not attaining statistical significance in this model) and unobserved heterogeneity (unmeasured factors which explain survival differences).

9. This pertains to both significant variables and variables without proven prognostic significance.

10. A particularly important variable is age. If the reference group encompasses only a limited age-range, the effect of age may easily be underestimated in extrapolations.

11. An observation is (right-) censored if no terminal event occured at the end of the follow-up. The occurence of censoring should not be related to the occurence of terminal events. Usually censoring is due to a short observation period. Other reasons are patient withdrawal, unrelated death, etc.
A particular application of censoring is in survival analysis of primary transplantations. If we treat retransplantation as terminal event we describe graft-survival; if we treat it as a censoring we describe patient survival of primary transplants; if we ignore the event we describe survival of transplanted patients, regardless of the number of retransplantations.

12. Note that statistical significance is about the *vertical* distance of two or more survival curves; clinical relevance is about the *horizontal* distance.

13. $\lambda(t) = f(t)/S(t)$

14. The reverse is not true.

15. $LY(\infty)$ of a health newborn usually is approximated with life tables resulting in about 74 year for males and about 80 year for females.

16. In statistical work $\lambda(t)$ frequently is preferred as it shows more clearly the changing pattern of the force of mortality (syn. hazard rate function).

17. Milani (1985) published a reduced sample regression analysis of survival cirrhotic patients. However, for unknown reasons he excluded all patients alive at the end of follow-up which makes the stated resemblance of his results with that of e.g. Christensen rather accidental.

18. For the actuarial method the usual (also in liver transplantation reports) 3-month interval was chosen.

19. This model seems more appropriate than the proportional hazards model as for biological reasons the duration of the effect of donor organ quality will be rather limited. Indeed we observed decreasing explanatory power of logistic regression models the later we defined the dependent variable, graft survival at x years.

20. This assumption requires the ratio of the hazard rates $\lambda_i(t)$ for i different levels of an independent variable to be constant. Proportionality may be inspected by plotting the log minus log survival function for i specific strata: $ln\ (-ln\ \hat{S}(t,z_i))$ versus t (see also Kalbfleisch, 1980; Hopkins, 1983). Tests are available (see e.g. Moreau, 1985; Lancaster, 1985).

21. This model requires two groups of otherwise comparable patients, one which is engaged in transplantation programme, and one that is not. From both groups data are drawn *at a comparable stage of their disease*, e.g. the occurrence of decompensated ascites. These data should cover the three types of variables mentioned: those which only influence prognosis without transplantation (also from the transplant group!), those which only influence prognosis with transplantation (also from the non-transplant group!), and those which affect both prognoses.

22. The time-dependent model as described by Christensen in 1986 treats each set of the observed values of prognostic variables of *one* patient at consecutive occasions (3 month-intervals) as separate cases in a time-fixed model (Christensen, 1986). Hence it is, strictly speaking, not a time-dependent model.

23. In computation of $\hat{S}(t)$ it may be convenient to approximate the stepfunction $\hat{\lambda}_{0k}(t)$, $\hat{\Lambda}_{0k}(t)$ or $\hat{S}_{0k}(t)$ by polynomial, Weibull or other functions.

24. For computational convenience $t_{max,j}$ may be rounded.

25. One of the main conclusions of this article is based on a common mistake in clinical survival analysis. Survival differences are explained by the occurrence of cholangitis in the course of survival. This is an obvious mistake: survival is stated to be predicted from birth onwards, hence the value of all covariates should be known at birth, the start of survival time (t_0).
    A similar error may frequently be found in analysis of survival after liver transplantation (Bonsel, 1988A). Operative blood loss (or an equivalent variable) is frequently included in the set of prognostic variables predicting survival after transplantation. As t_0 starts before surgery, "postdiction" rather than prediction is the case.
    These incorrect applications of survival techniques may have serious consequences when clinical policies are based on their outcomes.

26. Other reasons were (absolute numbers) cardiopulmonal contra-indication (3), severe osteoporosis (1), other surgical alternatives (5), miscellaneous (11).

27. For applications, see Buxton (1985) and especially the excellent analysis of Van Hout (1990).

Chapter 5

1. We think of the issue of length of life versus quality of life, care versus cure, priority setting based on health measures (survival, health status), etc.

2. For details see section 3.4.
In most specialties there exist classifications or descriptive systems which may serve for disease-specific health status measurement purposes (e.g. NYHA-classification, CCS-classification, AIMS classification).

3. One might even wish the application of the measurement instruments in the latter case to be as explicit and efficient as it is usually in the first case.

4. The empirical observations are not logically conferred to responses of the individual him/herself.

5. E.g. Andrews (1976) developed a refined classification of these concepts.

6. Berg (1973), Veit (1982) and Kane (1982) are still classics, which pay attention to valuation and description. They are rich of thoughtful considerations and little new has been put forward since.
With regard to description alone we could mention McDowell (1987), Spilker (1990), a recent supplement of Medical Care (1989, March) and the more technical and unconventional text of Streiner (1989). The evaluation of three important American instruments is interesting for its considerations (Read, 1987 and the discussion of Bergner, 1987).
De Haes (1985), though focussing on the cancer literature alone, demonstrates amongst other things the relevance of the distinction between objective versus the subjective (see *A4*) and elaborates on validity and design. Many of the theoretical considerations which can be found in her thesis on quality of life of cancer patients (De Haes, 1988) apply to decriptive measurement in general
The classics to our opinion reveal more of the conceptual considerations related to application on the macro level as opposed to the more recent works, which - understandably - focuss on the empirical work of the last decade.
The monograph of the Institute of Medicine (1989), though useful as a bibliography on the subject, also emphasizes applications rather than the theoretical considerations that may underpin the choices in each of the studies which is described. An extensive bibliography is published as a supplement of Medical Care (1990, December issue).

7. For a detailed treatment of the subject see for example Campbell (1976).

8. We distinguish the classical test theory, the latent trait theory and Gutman's theory on facet design. We agree with Dessens (1987) that operationalisation with a facet design as an addition to the classical approach would be worthwhile.

9. Based on the classical test theory.

10. A classification is defined as a measurement instrument with one or more items a, b, ..., p, each with a 1 - 1 correspondence to domains a, b, ..., p with a nominal or ordinal measurement level. An individual is classified as having a nominal or ordinal score on each of the domains.
A profile is defined as a measurement instrument with one or more items a_{1_n}, b_{1_n}, ..., p_{1_n}, each with a correspondence to domains a, b, ..., p by means of weighting and summing up of the empirical observations. Each individual has a score on each of the domains with a presumed interval or ratio measurement level.
An index finally, does not differ from the profile except for a final, usually weighted, summing up of all separate domain scores to one score, usually between 0 and 100. In this case an individual may be characterized by only one score with a presumed ratio measurement level.

11. The weighted sum of IPA and GWB yielding the Index of Well-Being does not add a new expression and hence is not regarded as a new expression.

12. We disregard other unidimensional parameters for their limited relevance, both from a theoretical and practical point of view. See also Drummond (1987 [p.15]).
The broader concept of outcome, which is adopted in QALY's and HYE's may include pain or disease avoided but also reassurance gained by information or emotional support and other process utilities (McGuire, 1988 [p.46-51,92,93,106]).

13. Utility as defined by definition 2. not necessarily complies with the requirements of definition 1 (Weber, 1987; Richardson, 1990 [p.12]) hence we may not regard it as a specific operationalisation within the general concept of definition 1.

14. Provisionally we assume that the sum of individual preferences equals societal preference (this might not be true as in some cases of vaccination).

15. This is not necessarily true: for good reasons it could consist of multiple branches and multiple levels of health status within a branch. To our knowledge there have been no attempts to use e.g. a two-branch health scenario. This is surprising in the case of the use of standard gambling as valuation method: a two-branch scenario is already included as calibration device.

16. Variation of the assumed duration only may provide insight into value of time. From our experience in a pilot study we conclude this requires extremely carefully defined calibration scenario's.

17. The willingness-to-pay method apparently is a surrogate for the first-best method, revealed preference. However, it is of little use for the insurmountable confounding effect of income and wealth. Moreover, we met a considerable number of adverse reactions of laymen respondents upon introduction of the method.
The equivalence method could be added, but we think this method does focus to much on the distribution problem - being aware this remains a problem.

18. Either an additive model or a multiplicative model.

19. Weighted by their probabilities or otherwise (McNeill, 1982)

20. *After* we established the feasibility of the EuroQol-system, we took a sample from health politicians (about 70). Apart from a higher response values were about the same.

Chapter 6

1. Technologies for diseases which are primarily defined by patient's complaints may give rise to empirical difficulties (e.g. CABG in angina pectoris).

2. From an economical point of view 'medically-defined need' does not exist as a separate entity. Scientific and political practice, in our view, justifies to speak of two viewpoints. The concepts of 'indication' and 'contra-indication' clearly refer to a medically defined need concept.

3. 'It can be defined either in terms of the type of illness or disability causing the need or of the treatment or facilities for treatment required to meet it.' We prefer to speak about 'and' instead of the second 'or'.

4. Hence there is a strong relationship between need estimation and patient outcome analysis.

5. Particularly in this case we should say 'should depend' due to obvious market imperfections.

6. In practice the first question to be adressed will regard alternative modes to solve the same medical problem.

7. A sovereign consumer choosing rationally the commodity that fits best his preference is not representative for the patient saving and receiving medical help. A supplier deciding to enter the market, calculating the optimal mix of quantity and quality of commodities (health care services) as to gain maximum profits usually is not representative for a health care provider.

8. Thus we accept a subjective feature of the real but unobservable quantity.

9. See the adjective 'acceptable' in the definition of Matthew.

10. These aggregation principles may be derived from the loss functions well-known in economic theory (maximin, minimax, maximax, expected utility). Different principles may yield entirely different results.

11. Prevention is the extreme case. But also heart transplantation for some diagnostic categories (arrythmia's).

12. Clinical language speaks of 'an indication to' rather than 'the need for' treatment.

13. In the ideal rational system this threshold does not depend on the disease involved.

14. A QALY/cost criterion may be chosen if a linear relation is accepted.

15. We think of case-control study, cohort study, sample survey, registry analysis (incidence, mortality, prevalence), mathematical modelling, case report, observational study, RCCT and expert opinion.

16. The assesment of need is an uncommon objective of medical research. If undertaken, it usually is 'piggyback' research, i.e. an investigational effort additional to an already existing design.

17. 'Detailed' refers to the distribution and the predictability of benefits.

18. Screening, as a complex of diagnostic, therapeutic and sometimes preventive, technologies is not regarded as a preventive technology in this classification.

19. They cannot be identified ex post either in primary prevention. Anonymity of the beneficiaries is a major source of the disrepute of prevention. The 1 in 10.000 who experience some adverse effect of prevention may succesfully blame the system; we shall experience the wrath of the 20 in 100.000 who die from preventable disease only at the other side.

20. E.g. minimal severity or absence of concurrent diseases.

21. Note that alcoholic cirrhoisis and hepatitis B were excluded from liver transplantation by the pilot programme protocol. These diseases probably will demonstrate geographical variation.

22. These favourable circumstances were a. the relevance of the mortality method combined with suitable mortality data, b. no competing treatments and no competing centres, and c. an excellent referral database.

Chapter 7

1. In fact 'utility' this way us a second best alternative: outcomes valued in currency units should be preferred.

2. Nothing so practical as excellent theory.

3. A tentative programme of MTA-methodologies needing further elaboration should include the following items. It is clear the nature of these methodological problems is quite divers. However, standardization (preferably at the international level) should be aimed at in all cases, despite the arbitrariness of some choices.
   The methodological issues are:
   1. non-experimental outcome analysis,
   2. quality-of-life measurement related to time,
   3. utility analysis,
   4. indirect costs,
   5. analysis of objective need,
   6. distributional analysis,
   7. ethical analysis.

4. The measurement of quality of care is but a straightforward extension of the MTA-concept (the reverse is also true).

5. We emphasize that consumer organisations have little to do with organisations representing the interests of patients with particular diseases. Sine nomen est omen.

6. Depending on the budgetting regime and the relation between the hospital management and the medical staff, economical expertise may be drawn from the hospital administrative staff or from outside.

7. This pressure curtailed MTA-activities from e.g. OTA and OHTA in the United States. Similar synergism may be expected in Europe, particular when 'Europe 1992' carries through.

8. Admittedly this is a paradox.

9. The erroneous expectation of impressive numbers reflecting scientific advance clearly originate from era of the succesful eradication of (most) infectious diseases (in Western countries). Mackenbach demonstrated the small number approach needed to demonstrate probably larger improvements really due to medical care in modern, and prosperous, hence healthy societies (Mackenbach, 1989).

REFERENCES

Andersen PK, Christensen E, Faureholdt L, Schlichting P. Measuring prognosis using the proportional hazards model. Scand J Stat 1983; 10: 49-52.

Anderson GM, Lomas J. Monitoring the diffusion of a technology; coronary artery bypass surgery in Ontario. Am J Publ H 1988; 78: 251-254.

Andrews FM. Withey SB. Social indicators of well-being. New York: Plenum Press, 1976. ISBN 0-306-30935-1.

Anonymous. Liver transplantation. Consensus conference. (Editorial). JAMA 1983A; 250: 2961-2964.

Anonymous. Hippocratic writings (Lloyd GER, editor). Middlesex: Penguin Books Ltd, 1983B. ISBN 0-14-044451-3.

Anonymous. National institutes of health consensus development conference statement: Liver transplantation-June 20-23, Hepatology 1984; 4: 107S-110S.

Anonymous. Inter-regional variations in the utilization of hospital services (in Dutch). Final report of the Task Force of the Royal Dutch College of Physicians on the Inter-regional variations in the utilization of hospital services. Utrecht: Royal Dutch College of Physicians, 1989. ISBN 90-71994-03-1.

Anonymous. The Steering Group on Future Health Scenario's: structure and task (in Dutch). Rijswijk: Steering Group on Future Health Scenario's, 1987.

Anonymous. Declaration of the Government on the bounderies of health care (in Dutch). Papers of the Parliament 1987-1988, No. 20620. Den Haag: Staatsuitgeverij, 1988.

Banta HD, Kemp K (eds). The management of health care technology in nine countries. New York: Springer Publishing Company, 1982.

Banta HD. The uses of modern technologies: problems and perspectives for industrialized and developing countries. Bull Pan Amer Health Organ 1984; 18: 139-150.

Banta HD. Pushing the limits: health care technology assessment. Inaugural adress. Utrecht: Elinkwijk, 1990.

Banta HD, Geleijns A. Anticipating and assessing health care technology. Volume 1. General considerations and policy conclusions. Dordrecht: Martinus Nijhoff Publishers, 1987. ISBN 0-89838-897-X.

Beauchamp TL, Childress JF. Principles of biomedical ethics. Oxford: Oxford University Press, 1989. ISBN 0-19-505-9026.

Beaufort ID de, Dupuis HM. Guide to medical ethics (in Dutch). Assen: Van Gorcum, 1988. ISBN 90-232-2377-2.

Beeck EF van, Mackenbach JP, Oortmarssen GJ van, Barendregt J, Habbema JDF, Maas PJ van der. Scenarios for the future development of accident mortality in The Netherlands. Health Policy 1989; 11: 1-17.

Benedetti J, Yuen K, Young L. Life-tables and survival functions. In: Dixon WJ (ed). BMDP Statistical Software, 1983 printing with additions. Los Angeles: University of California Press, 1983. ISBN: 0-520-04408-8.

Berg RL (eds). Health status indexes. Proceedings of a conference conducted by Health Services Reseacrh Tucson, Arizona, October 1-4, 1972. Chicago: Hospital Research and Educational Trust, 1973. ISBN 0-87914-029-9.

Bergner M, Kaplan RM, Ware JE. Evaluating health measures. Commentary: Measuring overall health: an evaluation of three important appraches. J Chron Dis 1987; 40: 23S-26S.

Blackorby C, Donaldson D. Utility vs equity. Some plausible quasi-orderings. J Publ Econ 1977; 7: 365-381.

Bonsel GJ. Survival gain after liver transplantation (in Dutch). In: A new liver; a decade of liver transplantations in Groningen (Over lever en 10 jaar levertransplantaties in Groningen; in Dutch). Pruim J (ed). Groningen: Academisch Ziekenhuis Groningen, 1990A. ISBN 90-5252-005-4.

Bonsel GJ. The costs and effects of heart transplantation. Volume 2B: Quality of life before and after heart transplantation - Interview (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988B. ISBN: 90-72245-27-X.

Bonsel GJ. The costs and effects of heart transplantation. Volume 1: A summary and comparison of two foreign technology assessments of heart transplantation (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988A. ISBN 90-72245-10-5.

Bonsel GJ, Bot ML, Veer F van 't. The costs and effects of liver transplantation. Volume 7: The need for liver transplantation (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988G. ISBN: 90-72245-40-7.

Bonsel GJ, Bot ML, Boterblom A, Veer F van 't. The costs and effects of heart transplantation. Volume 2C: Quality of life before and after heart transplantation - Results (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988C. ISBN: 90-72245-15-6.

Bonsel GJ. The costs and effects of liver transplantation. Volume 3A: description of the liver transplantation programme - data collection forms (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988D. ISBN: 90-72245-31-8.

Bonsel GJ, Veer F van 't, Bot ML. The costs and effects of liver transplantation. Volume 3B: description of the liver transplantation programme - the Groningen experience 1977 - 1987 (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988E. ISBN 90-72245-27-2.

Bonsel GJ, Veer F van 't, Bot ML. Costs and effects of liver transplantation. Volume 4: Prognosis of liver transplantation, the Groningen program 1977 - 1987 (in Dutch). Rotterdam: Institute of Public Health and Social Medicine, Erasmus University Rotterdam, 1988F. ISBN: 90-72245-39-3.

Bonsel GJ, Habbema JDF, Bot ML, Veer F van 't, Charro FT de, Maas PJ van der. A technology assessment of liver transplantation. The liver transplantation programme in Groningen 1977 - 1987 (in Dutch). Ned Tijdschr Geneeskd 1989B; 133: 1406-1414.

Bonsel GJ, Erdman RAM, Mast RC van der, Balk AHMM, Maas PJ van der. Psychosocial aspects of heart transplantation (in Dutch). Ned Tijdschr Geneesk 1990B; 134: 227-231.

Bonsel GJ, Haes JCJM de. Quality of life (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Bot ML, Bonsel GJ, Berg H van de. The costs and effects of heart transplantation. Volume 3: The history of heart transplantation (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988B. ISBN: 90-72245-14-8.

Bot ML, Bonsel GJ, Veer F van 't. The costs and effects of heart transplantation. Volume 8: The need for heart transplantation (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988A. ISBN: 90-72245-16-4.

Bot ML, Bonsel GJ, Maas PJ van der. Valuation of health states by the general public: feasibility of a standardized measurement procedure. Soc Sci Med 1990; 31: 1201-1206.

Bradshaw J. A taxonomy of social need. In: McLachlan G (ed). Problems and progress in medical care, 7th series, London: Oxford University Press, 1972.

Breslow N. Analysis of survival data under the proportional hazards model. Int Stat Rev 1975; 43: 45-58.

Brock DW, Wartman SA. When competent patients make irrational choices. N Eng J Med 1990; 322: 1595-1599.

Buxton M, Acheson R, Caine N, Gibson S, O'Brien B. Cost and benefits of the heart transplant programmes at Harefield and Papworth Hospitals: Final Report. Department of Economics, Brunel University and Department of Community Medicine, University of Cambridge, 1985.

Callahan D. Rationing medical progress. The way to affordable health care. N Eng J Med 1990; 322: 1810-1813.

Campbell A, Converse PE, Rodgers WL. The quality of American life. New York: Russel Sage Foundation, 1976. ISBN 0-87154-194-7.

Campbell DT, Stanley JC. Experimental and quasi-experimental designs for research. Boston: Houghton Mifflin Company, 1966. ISBN 0-395-30787-2.

Casparie AF. The use of formal group methods in medical technology assessment (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Casparie AF. Decision-making on medical technologies: health policies and medical practice (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Charro FT de, Ineveld BM van. Cost-effectiveness-analysis (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989A. ISBN 90-6016-229-3.

Charro FT de, Banta, HD. Transplant policy in the Netherlands. Int J Techn Assessment Health Care 1986; 2: 533-544.

Charro FTH de, Bonsel GJ, Hout BA van. The costs and effects of heart transplantation. Volume 9: Final report (in Dutch). Rotterdam: Department of Economics, Faculty of Law & Institute of Public Health, Erasmus University Rotterdam, 1988. ISBN: 90-72245-19-9.

Charro FT de. The cost-effectiveness of treatment of end-stage renal disease (in Dutch). Thesis. Delft: Oberon, 1989B. ISBN 90-5166-018-9.

Child CG, Turcotte JG. Surgery and portal hypertension. In: Child CG (ed). The liver and portal hypertension. Philadelphia: Saunders Company, 1968.

Christensen E, Krintel JJ, Hansen SM, Johansen JK, Juhl E. Prognosis afyer the first episode of gastrointestinal bleeding or coma in cirrhosis. Scand J Gastroent 1989; 24: 999-1006.

Christensen E, Multivariate survival analysis using Cox's regression model, Hepatology 1987; 7: 1346-1358.

Christensen E, Neuberger J, Crowe J, Altman DG, Popper H, et al. Beneficial effects of azathioprine and prediction of prognosis in primary biliary cirrhosis. Final results of an international trial. Gastroenterology 1985; 89: 1084-1091.

Christensen E. Individual therapy-dependent prognosis based on data from controlled clinical trials in chronic liver disease. Dan Med Bull 1988; 35: 167-182.

Christensen E, Schlichting P, Faureholdt L, Gluud C, Andersen PK, et al. Prognostic value of Child-Turcotte criteria in medicallly treated cirrhosis. Hepatology 1984; 4: 430-435.

Christensen E, Schlichting P, Andersen PK, Fauerholdt L, Schou G, et al. Updating prognosis and therapeutic effect evaluation in cirrhosis with Cox's multiple regression model for time-dependent variables. Scand J Gastroenterol 1986; 21: 163-174.

Cochrane AL. Effectiveness and efficiency: random reflections on health services. London: Nuffield Provincial Hospitals Trust, 1972.

Fahrenfort M. A doctor novel. An explorative investigation of the self-perception of patients and doctors (in Dutch). Thesis. Amsterdam: Free University Press, 1985. ISBN 90-6256-135-7.

Conn HO. A peek at the Child-Turcotte classification. (Editorial). Hepatology 1981; 1: 673-676.

Cox DR. Regression models and life tables. J Royal Stat Soc 1972; 34: 187-220.

Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall, 1984. ISBN 0-412-24490-X.

Cutler SJ, Ederer F. Maximum utilization of the life table method in analyzing survival. J Chron Dis 1958; 8: 699-712.

Dessens JAG, Janssen W. From theoretical concept to empirical variable: tradition and criticism. (Operationaliseren: traditie en kritiek; in Dutch). Thesis. Utrecht: Elinkwijk, 1987.

Dickson ER, Grambsch PM, Fleming TR, Fisher LD, Langworthy A. Prognosis in primary biliary cirrhosis: model for decision making. Hepatology 1989; 10: 1-7.

Drummond MF, Stoddart GL, Labelle R, Cushman R. Health economics: an introduction for clinicians. Ann Int Med 1987B; 107: 88-92.

Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1987A. ISBN 0-19-261601-3

Eddy DM. Screening for cancer: theory, analysis and design. Englewood Cliffs (New Jersey): Prentice Hall Inc., 1980. ISBN 0-13-796789-6.

Essink-Bot ML, Agt HME van, Bonsel GJ. NHP or SIP? A comparative analysis among chronic diseased (in Dutch). Tijdschr Soc Gen (in press) 1991.

Euroqol-studygroup. EuroQol, a new facility for the measurement of health related quality-of-life. Health Policy 1990, 16: 199-208.

Evans RW, Manninen DL, overcast TD, Garrison LP, Yagi J, Merrikin K, Jonsen AR, et al. The national heart transplantation study: final report. Battelle Human Affairs Research Centers, Seattle, WA, 1984.

Farber SJ. On the necessity of a scientific base for choice. In: Banta HD (ed). Resources for health. Technology assessment for policy making. New York: Prager, 1982. ISBN 0-03-061341-8., 1980

Foucault MF. The birth of the clinic. New York: Random House Inc., 1975. ISBN 0-394-71097-5.

Foucault MF. Microphysik der Macht. Uber Strafjustiz, Psychiatrie und Medizin. Berlin: Merve Verlag GmbH, 1976.

Foucault MF. The history of sexuality. Volume 1: an introduction. New York: Random House Inc., 1980. ISBN 0-394-74026-2.

Freidson E. Profession of medicine (1972). (De medische professie; translated in Dutch). Lochem: De Tijdstroom, 1981. ISBN 90-6087-259-2.

Froberg DG, Kane RL. Methodology for measuring health state preferences I-IV. J Clin Epid 1989; 42; 345--354; 459-471; 585-592; 675-685.

Fuchs VR. Who shall live? New York: Basic Books, 1983. ISBN 0-465-09186-5.

Garceau AL, Chalmers TC. The Boston Inter-Hospital Liver Group. The natural history of cirrhosis. N Eng J Med 1963; 268: 469-473.

Gehan EA. A generalized Wilcoxon-test for comparing arbitrarily single-censored samples. Biometrika 1965; 52: 203-223.

Geleijns AC, Rigter H. Health care technology assessment in the Netherlands. Int J Techn Assessment in Health Care 1990; 6: 157-174.

Gilks WR, Gore SM, Bradley BA. Analyzing transplant survival data. Transplantation 1986; 42: 46-49.

Ginès P, Quintero E, Arroyo V, Ters J, Bruguera M, Rimola A. Compensated cirrhosis: natural history and prognostic factors. Hepatology 1987; 7: 122-128.

Ginès P, Titó L, Arroyo V, Planas R, Panés J, et al. Randomized comparative study of therapeutic paracentesis with and without intravenous albumin in cirrhosis. Gastroenterology 1988; 94: 1493-1502.

Glass GV, McGaw B, Smith ML. Meta-analysis in social research. Beverly Hills: Sage 1981.

Grambsch PM, Dickson ER, Kaplna M, LeSage G, Fleming TR, Langworthy AL. Extramural cross-validation of the Mayo primary biliary cirrhosis survival model establishes its generalizability. Hepatology 1989; 10: 846-850.

Greer AL. Advances in the study of innovations in health care organisations. Milb Memorial Fund Quarterly 1977; 55: 505-532.

Gudex C, Kind P. The QALY Toolkit. Centre for Health Ecnomics Discussion Paper 39. York: University of York, 1985.

Haagsma EB. Orthotopic liver transplantation. Clinical and laboratory results. Thesis. Groningen: Dijkhuizen van Zanten: 1990.

Habbema JDF, Lubbe JTN, Agt HME van, Ballegooijen M van, Koopmanschap MA, Oortmarsen GJ. The costs and effects of population screening for cervival cancer. Final Report (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988A. ISBN 90-72245-33-4.

Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Habbema JDF, Bonsel GJ. The costs and effects of liver transplantation. Volume 1: final report. The Dutch liver transplantation programme 1977-1987 (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1988. ISBN: 90-72245-28-8.

Habermas J, Technik und Wissenschaft als Ideologie (1968). Translated in Dutch in: Baars J. A selection from the works of Jürgen Habermas. Deventer: Van Loghum Slaterus, 1976. ISBN 90-6001-407-3.

Haes, JCJM de, Knippenberg FCE van. The quality of life of cancer patients: a review of the literature. Soc Sci Med 1985; 20: 809-817.

Haes, JCJM de. Quality of life of cancer patients (in Dutch). Thesis. Amsterdam: Swets & Zeitlinger, 1988.

Hällèn J, Krook H. Follow-up studies on an unselected 10-year material of 360 patients with liver cirrhosis in one community. Acta Med Scand 1963; 25: 43-55.

Harris J. QALYfying the value of life. J Med Ethics 1987; 13: 117-123.

Harrison MR. Organ procurement for children: the anencephalic fetus as donor. Lancet 1986; ii: 1383-1386.

Harsanyi JC. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. J Political Econ 1955; 63: 309-321.

Have H ten. Ethics and economics in health care: a medical philosopher's view. In: Mooney GH, McGuire AJ. Medical ethics and economics in health care. Oxford: Oxford University Press, 1988. ISBN 0-19-261672-2.

Health Council (Gezondheidsraad). Liver transplantation (in Dutch). 's-Gravenhage: Health Council (Gezondheidsraad), 1989.

Health Council (Gezondheidsraad). On brain death guidelines. (in Dutch). 's-Gravenhage: Health Council (Gezondheidsraad), 1983.

Health Council (Gezondheidsraad). General considerations on organ transplantation (in Dutch). 's-Gravenhage: Health Council (Gezondheidsraad), 1987.

Health Council (Gezondheidsraad). On the boundaries of health care (in Dutch). 's-Gravenhage: Health Council (Gezondheidsraad), 1986.

Health Council (Gezondheidsraad). Heart transplantation (in Dutch). 's-Gravenhage: Health Council (Gezondheidsraad), 1988.

Hoffenberg R. Modern medecine: prospects and problems. In: Austin JM (ed). New prospects for medecine. Oxford: Oxford University Press, 1988. ISBN 0-19-261647.

Holland WW, Ipsen J, Kostrzewski J (eds). Measurement of levels of health. WHO Regional Publications European Series No. 7. Copenhagen: WHO, 1979. ISBN 92-9020-107-X.

Hoogeveen RT, Brouwer JJLM. Scenario-analysis in public health (in Dutch). Rijswijk: Steering Group on Future Health Scenario's, 1989. ISBN 90-6224-247-2.

Hop WCJ, Hermans J. Statistical analysis of survival times (in Dutch). Tijdschr Soc Geneesk 1981; 59: 279-288.

Hopkins A. Survival analysis with covariates - Cox models. In: Dixon WJ (ed). BMDP Statistical Software, 1983 printing with additions. Los Angeles: University of California Press, 1983. ISBN 0-520-04408-8.

Hout BA van, Habbema JDF, Bonsel GJ. The costs and effects of heart transplantation. Volume 5: The prognosis of heart transplantation (in Dutch). Rotterdam: Department of Economics, Faculty of Law & Institute of Public Health, Erasmus University Rotterdam, 1988. ISBN: 90-72245-10-5.

Hout BA. Heart transplantation. Costs, prognosis, future impact (in Dutch). Thesis. Delft: Oberon, 1990. ISBN 90-5166-135-5.

Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat in Medicine 1990; 9: 1303-1325.

Houwelingen JC van, Stijnen T. Meta-analysis (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Houwen RJ, Zwierstra RP, Severijnen RSVM, Bouquet J, Madern G, Vos A. Prognosis of extrahepatic biliary atresia. Arch Dis Child 1989; 64: 214-218.

Iglehart JK. Japan's medical care system - part two. N Eng J Med 1988; 319: 1166-1172.

Illich I. Medical Nemesis (1974). (Het Medisch Bedrijf; in Dutch). Baarn: Het Wereldvenster, 1975. ISBN 90-293-0242-9.

Institute of Medecine. Assessing medical technologies. Washington DC: National Academy Press, 1985. ISBN 0-309-03583-X.

Institute of Medecine. Quality of life and technology assessment. Washington DC: National Academy Press, 1989. ISBN 0-309-04098-1.

Jennett B. High technology medecine. Benefits and burdens. Oxford: Oxford University Press, 1986. ISBN 0-19-261588-2.

Ji-yao W, Hou-yu L, Wu-nan Z, Cai-Yun D, et al. An analysis of prognostic factors in cirrhosis. Chin Med J 1989; 102: 266-272.

Jones-Lee MW. The value of life: an economic analysis. London: Martin Robertson, 1976.

Jong de G, Rutten FFH. Justice and health for all. Soc Sc Med 1983; 17: 1085-1095.

Jong KP de, Blankensteijn JD, Hesselink EJ, Laméris JS, Terpstra OT. Partial hepatectomy in benign and malignant liver disease; experience in 94 patients (in Dutch). Ned Tijdschr Geneeskd 1989; 133: 2385-2388.

Jönsson B, Rutten FFH, Vang J (eds). Policymaking in health care. Changing goals and new tools. Linköping: Samhall Klintland - Linköping Collaborating Center, Linköping University, Sweden, 1990. ISBN 91-7870-693-9.

Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley & Sons, 1980. ISBN 0-471-05519-0.

Kane RA, Kane RL (eds). Values and long-term care. Lexington: RAND Corporation, DC Heath and Company, 1982.

Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. JASA 1958; 53: 457-481.

Kasiske BL, Neylan JF, Riggio RR, Danovitch GM, Kahana L, Alexander SR, White MG. The effect of race on acces and outcome in transplantation. N Eng J Med 1991; 324: 302-307.

Keeler EB, Cretin S. Uses of cost-benefit-analysis. J Health Econ 1987; 6: 275-278.

Kind P. The design and construction of quality of life measures. Centre fro Health Ecnomics Discussion Paper 43. York: University of York, 1986.

Kirschner B, Guyatt G. A methodological framework for assessing health indices. J Chron Dis 1985; 38: 27-36.

Kleinbaum DG, Kupper LL, Morgenstern H. Epidemiologic research. Principles and quantitative methods. New York: Van Nostrand Reinhold, 1982. ISBN 0-534-97950-5.

Knippenberg FCE, Haes JCJM. Measuring the quality of life of cancer patients: psychometric properties of instruments. J Clin Epid 1988; 41: 1043-1053.

Koopmanschap MA, Roijen L van, Bonneux L. Costs of illness in the Netherlands (in Dutch). Rotterdam: Institute of Public Health, Erasmus University Rotterdam, 1991. ISBN 90-72245-54-7.

Kroon ML de, Bonsel GJ. Future health scenarios in obstetrics, gynaecology and reproductive medecine. A feasibility study (in Dutch). Rotterdam: Institute for Medical Technology Assessment, Erasmus University Rotterdam, 1990. ISBN 90-72245-2.

Krüger H. Economic analysis of solid organ transplantation: a review for policy makers. Health Policy 1989; 13: 1-17.

Lancaster T, Chesher A. Residuals, tests, plots with job matching illustration. Annales de l'INSEE 1985; nr. 59/60: 47-69.

Lammerts van Bueren WM. Dealing with uncertainty. The quantification of uncertainties in decision-making (in Dutch). Den Haag: Staatsuitgeverij, 1987. ISBN 90-12-05568-7.

Leaf A. Cost-effectiveness as a criterion for Medicare coverage. N Eng J Med 1989; 321: 898-900.

Lee E. Statistical methods for survival data analysis. Belmont (California): Lifetime Learning Publications, 1980. ISBN 0-534-97987-4.

Leeuwen E van, Feenstra L. Ethics, medicine and health care; the Dutch Compendium on Medical Ethics. Ned Tijdschr Geneesk 1989; 133: 1370-1374.

Levinsky NG. Age as a criterion for rationing health care. N Eng J Med 1990; 322: 1813-1816.

Lipscomb J. Value preference for health: meaning, measurement , and use in programme evaluation. In: Vlaues and long-term care. KaneRL, Kane RA, (eds). Lexington: D.C.Heath and Company, 1982.

Llach J, Ginès P, Arroyo V, Rimola A, Titó L, et al. Prognostic value of arterial pressure, endogenous vasoactive systems, and renal function in cirrhotic patients admitted to the hospital for the treatment of ascites. Gastroenterology 1988; 94: 482-487.

Luce BR, Elixhauser A. Standards for the socioeconomic evaluation of health care services. New York: Springer Verlag, 1990. ISBN 0-387-52264-6.

Maas PJ van der. Aging and public health. In: Health and aging. Schroots JJF, Birren JE, Svanborg A (eds). New York: Springer Publishing Company, 1988. ISBN 90-265-0866-2.

Maas PJ van der, Habbema JDF. The development of mathematical models and the availability of information: two preliminaires to health policy (in Dutch). In: Health policy on the boundaries of health care. Dekker E, Wijnberg B (eds). Alphen a/d Rijn: Samson Stafleu, 1986. ISBN 90-6016-784-8.

Maas PJ van der. They lived happily ever after. (Lang zullen we leven; in Dutch). Annual inaugural adress of the Erasmus University Rotterdam. Rotterdam: Department of Public Health and Social Medicine, Erasmus University Rotterdam, 1989. ISBN 90-72245-48-2.

Mackenbach JP. Geographical variations in some highly prevalent surgical interventions, 1985 (in Dutch). Ned Tijdschr Geneesk 1990; 134: 953-957.

Mackenbach, JP. Mortality and medical care. Thesis. Rotterdam, 1989.

Mantel N. Evaluation of survival data and two new rank order statistics arising in tis consideration. Cancer Chemotherapy Reports 1966; 50: 163-170.

Manton KG, Stallard. Chronic disease modeling: measurement and evaluation of the risks of chronic disease processes. New York: Oxford University Press, 1988. ISBN 0-85264-286-5.

Markus BH, Dickson ER, Grambsch PM, Fleming TR, Mazzaferro V, et al. Efficacy of liver transplantation in patients with primary biliary cirrhosis. N Engl J Med 1989; 320: 1709-1713.

Mathew, 1971 cited in Williams, 1978 [p.32].

McDowell I, Newell C. Measuring health: a guide to rating scales and questionnaires. New York: Oxford University Press, 1987. ISBN 0-19-504101-1.

McGuire AJ, Henderson JB, Mooney GH. The economics of health care. An introductory text. London: Routledge & Kegan Paul, 1988. ISBN 0-7102-1300-X.

McKeown T, The role of medecine: dream, mirage or nemesis. London: Nuffield Provincial Hospitals Trust, 1976. ISBN 0-900574-24-0.

McNeill BJ, Pauker SG. Incorporation of patient values in medical decision making. In: Critical issues in medical technology. McNeill BJ, Cranatho EG (eds). Boston: Auburn House Publishing Company, 1982.

McNeill BJ, Pauker SG. Optimizing patient and societal decision making by the incorporation of individual values. Chapter 8 in: Values and long-term care. Kane RL and Kane RA (eds). Lexington: RAND Corporation, DC Heath and Company, 1982.

McPherson K. International differences in medical care practices. In: Health care systems in transition. The search for efficiency. OECD Social Policy Studies Number 7. Paris: OECD, 1990.

Mehrez A, Gafni A. Evaluating health-related quality of life: an indifference curve interpretation for the time trade-off technique. CHEPA Working Paper Series #13. Toronto: McMaster University, Hamilton, Ontario, 1989B.

Mehrez A, Gafni A. Quality adjusted life-years, utility theory, and healthy-years equivalents. Med Dec Making 1989; 9A: 142-149.

Milani A, Marra L, Siciliano M, Rossi L. Prognostic significance of clinical and laboratory parameters in liver cirrhosis. A multivariate statistical approach. Hepato-gastroenterology 1985; 32: 270-272.

Milholland AV, Wheeler SG, Heieck JJ. Medical assessment by a Delphi group opinion technique. N Eng J Med 1973; 288: 1272-1275.

Mintzberg H. The structuring of organisations. Englewood Cliffs (New Jersey): Prentice Hall Inc., 1979. ISBN 0-13-855270-3.

Mishan EJ. Cost-benefit analysis. London: George Allen & Unwin, 1984. ISBN 0-04-338099-9.

Mooney G. Economocs, medicine and health care. Brighton: Wheatsheaf Books Ltd., 1986. ISBN 0-7450-0035-5.

Mooney GH. The valuation of human life. London: Macmillan, 1977.

Moreau T, O'Quigley J, Mesbah. A global goodness-of-fit statistic for the proportional hazards model. Applied Statist 1985; 34: 212-218.

Mosteller F. Innovation and evaluation. Science 1981; 211: 881-886.

Musschenga AW, Borst-Eilers E. How to achieve equity (in Dutch). Medisch Contact 1987; 42: 13-19.

Neuberger J, Altman DG, Polson R, Buckels J, Rolles K, et al. Prognosis after liver transplantation for primary biliary cirrhosis. Transplantation 1989; 48: 444-447.

Neuberger J, Altman DG, Christensen E, Tygstrup N, Williams R. Use of a prognostic index in evaluation of liver transplantation for primary biliary cirrhosis. Transplantation 1986; 41: 713-716.

Neuhauser D, Lewicki AM. What do we gain from the sixth stool guaiac? N Eng J Med 1975; 293: 226-228.

NHIB (Ziekenfondsraad). Liver transplantation. Advice to the Secretary of Health on the reimbursement of liver transplantation (in Dutch). NHIB Publications No. 429. Amstelveen: Ziekenfondsraad, 1988.

NHIB (Ziekenfondsraad). The final recommendations on the extension of national health insurance scheme (in Dutch). NHIB (Ziekenfondsraad) Publications No. 319. Amstelveen: Ziekenfondsraad, 1986.

NHIB (Ziekenfondsraad). The interim report on the extension of national health insurance scheme (in Dutch). NHIB (Ziekenfondsraad) Publications No. 247. Amstelveen: Ziekenfondsraad, 1983.

NHIB (Ziekenfondsraad). Evaluation of Medical Technology Assessments. Proceedings of NHIB-symposium, Amstelveen 31-01-1989 (in Dutch). Amstelveen: Ziekenfondsraad, 1989.

O'Brien B. An overview of the estimation and evaluation of clinical risks. In: Mann RD (ed). Risk and consent to risk in medicine. Camforth, Lancs (UK): Parthenon Publishing Group, 1989. ISBN 1-85070-263-2.

O'Grady JG, Williams R. Present position of liver transplantation and its impact on hepatological practice. Gut 1988; 29:566-570.

O'Quigley J. Asymptomatic variance formulae for some measures used as prognostic indices. Statistician 1983; 32: 351-352.

Office of Technology Assessment (OTA). Background Paper #4: The Management of Health Care Technology in ten Countries, Washington DC: OTA, 1980.

Olsen SA. Group planning and problem solving methods in engineering management. New York: John Wiley, 1982.

Oortmarssen GJ, Habbema JDF, Maase PJ van der, Koning HJ de, Collette HJA, Verbeek ALM, Geerts AT, Lubbe KTN. A Model for breast cancer screening. Cancer 1990; 66: 1601-1612.

Pauker SG, Kassirer JP. Decision analysis. N Eng J Med 1987; 316: 250-258.

Peto R, Peto J. Asymptotically efficient rank invariant procedures. J Royal Stat Soc 1972; 135: 185-207.

Phelps CE, Parente ST. Priority setting in medical technology and medical practice assessment. Med Care 1990; 28: 703-723.

Pliskin JS, Shephard DS, Weinstein MC. Utility functions for life years and health status. Op Research 1980; 28: 206-224.

Porayko MK, Wiesner RH, LaRusso NF, Ludwig J, MacCarty RL, et al. Patients with asymptomatic primary sclerosing cholangitis frequently have progressive disease. Gastroenterology 1990; 98: 1594-1602.

Powell WJ, Klatskin G. Duration of survival in patients with Laennec's cirrhosis. Am J Med 1968; 44: 406-420.

Pruim J, Bonsel GJ, Klompmaker IJ, Veer F van't, IJ Klompmaker, Vos RD, Slooff MJH. Donor parameters and survival of recipients liver grafting. A statistical analysis. (Abstract) 4th International Symposion on Organ Procurement and Preservation, Minneapolis, september 1989.

Ratnoff OD, Patek AJ Jr. The natural history of Laennec's cirrhosis of the liver. Medicine 1942; 21: 207-268.

Rawls J, A theory of justice. Oxford: Oxford University Press, 1972. ISBN 0-19-881301-5.

Reiser SJ. Medecine and the reign of technology. Cambridge: Cambridge University Press, 1978. ISBN 0-521-21907-8 / 0-521-28223-3.

Relman AS. Assessment and accountability. N Eng J Med 1988; 319: 1220-1222.

Richardson J. Cost-utility analysis: what should be measured - utility, value or healthy years equivalents. Paper presented to the 2nd World Congress on Health Economics, Zurich, Switzerland, 1990

Rigter H. The social look of medecine. On medical technologies in the 20th century. (Het sociale gezicht van de geneeskunde. Over de ontwikkeling van medische technologie in deze eeuw; in Dutch). Inaugural adress. Rotterdam: Institute for Medical Technology Assessment, Erasmus University Rotterdam, 1989.

Ringe B, Wittekind C, Bechstein WO, Bunzendahl H, Pichlmayr R. Th eorle of liver transplantation in hepatobiliary malignancy. Ann Surg 1989; 209: 88-98.

Roos BR. The economic evaluation of medical technologies and the valuation of future medical costs (in Dutch). Research Paper. Rotterdam: Department of Health Policy and Management, Erasmus University Rotterdam, 1990.

Roscam Abbing H. Legal aspects in MTA. An introduction. (Gezondheidsrecht, een terreinverkenning. In Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Rossum W van. Medical technology assessment. An analysis of the perceived value of the Dutch technology assessments of heart and liver transplantation, and in-vitro-fertilisation (in Dutch). Medisch Contact 1990;45:509-511.

Rutten FFH. Issues in cost-analysis (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Senn SJ, Auclair P. The gaphical representation of clinical trials with particular reference to measurements over time. Stat in Medicine 1990; 9: 1287-1302.

Showstack JA, Schroeder SA, Matsumoto MF. Changes in the use of medical technologies, 1972-1977. N Eng J Med 1982; 306: 706-712.

Siegel JH, Williams. A computer based index for the prediction of operative survival in patients with cirrhosis and hypertension. Ann Surg 1969; 169: 191-210.

Slooff MJH, Bams JL, Sluiter WJ, Klompmaker IJ, Hesselink EJ, Verwer R. A modified cannulation technique for veno-venous bypass during oxthotopic liver transplantation. Transplantation Proc 1989; 21: 2328-2329.

Spilker B (ed). Quality of life assessments in clinical trials. New York: Raven Press, 1990.

Stevenson LW, Fowler MB, Schroeder JS, Dracup KA, Clark SH. Patients denied cardiac tranplantation for non-medical criteria: a control group. J Am Coll Card 1986; 7: 9A.

Stevenson LW, Fowler MB, Schroeder JS, Stevenson WG, Dracup KA, Fond V. Poor survival for patients with idiopathic cardiomyopathy considered too well for transplnatation. Am J Med 1987; 83: 871-876.

Streiner DL, Norman GR. Health measurement scales. A practical guide to their development and use. Oxford: Oxford University Press, 1989. ISBN 0-19-261773-7.

Swaan A de. Essays 1971-1981. (De mens is de mens een zorg. Opstellen 1971-1981; in Dutch). Amsterdam: Meulenhoff, 1984. ISBN 90-290-1138-6.

Torrance GW. Measurement of health utilities for economic appraisal. A review. J Health Econ 1986; 5: 1-30.

Truin GJ. A computer simulation model of dental care (in Dutch). Thesis. Oss (The Netherlands): Witsiers, 1982.

Turnbull BW, Brown BW, Hu M. Survivorship analysis of hearttransplant data. J Am Stat Assoc 1974; 69: 74-79.

Vandenbroucke JP. The National Office of Statistics: the cemetery of deatch certificates (Het Centraal Bureau voor de Statistiek: de begraafplaats van onze doodsoorzaken; in Dutch). Ned Tijdschr Geneesk 1989; 133: 2112-2114.

Veit CT, Rose BJ, Ware JE. Effets of physical and mental health on health -state preferences. Med Care 1982; 20: 386-401.

Wachter MAM de. Ethical aspects (in Dutch). In: Habbema JDF, Casparie AF, Mulder JH, Rutten FFH (eds). Medical technology assessment and health policy. Alphen aan den Rijn: Samson Stafleu, 1989. ISBN 90-6016-229-3.

Wagstaff A, Doorslaer EKA van, Paci P. Equity in the finance and delivery of health care: some tentative cross-country comparisons. Oxford Review of Economic Policy 1989; 5: 89-112.

Weber M, Camerer C. Recent developments in modelling preferences under risk. Op Research Spectrum 1987; 9: 129-151.

Weisbrod BA, Test MA, Stein LI. Alternative to mental hospital treatment. II. Economic benefit-cost analysis. Arch Gen Psychiatry 1980 ;37 :400-405.

Wennberg JE, Gittelsohn A. Small area variations in health care delivery: a population based health infromation system can guide planning and regulatory decision making. Science 1973; 182: 1102-1109.

White KL. Political arithmetick: integration of CEA into health systems. In: Banta HD (ed). Resources for health. Technology assessment for policy making. New York: Prager, 1982. ISBN 0-03-061341-8.

Wiesner RH, Grambsch PM, Dickson ER, Ludwig J, MacCarty RL, et al. Primary sclerosing cholangitis: natural history, prognostic factors adn survival analysis. Hepatology 1989; 10: 430-436.

Williams A. Need - an economic exegesis. In: Culyer AJ, Wright KG (eds). Economic aspects of health services. Oxford: Martin Robinson, 1978. ISBN 0-85520-227-0.

World Health Organization. Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death. 1975 Revision. Volume 1. Geneva, 1977.

APPENDIX

Orthotopic liver transplantation in The Netherlands. The results and impact of a medical technology assessment

Gouke J. Bonsel
Marie-Louise Essink-Bot
Frank T. de Charro
Paul J. van der Maas
J. Dik F. Habbema

Health Policy, 16 (1990) 147-161
Published with permission of the editor

Orthotopic liver transplantation in The Netherlands. The results and impact of a medical technology assessment

Gouke J. Bonsel[1,2], Marie-Louise Essink-Bot[1], Frank T. de Charro[3], Paul J. van der Maas[1] and J. Dik F. Habbema[1]

[1]Institute for Medical Technology Assessment, Medical Faculty, [2]Department of Public Health and Social Medicine and [3]Department of Economics, Faculty of Law, Erasmus University, Rotterdam, The Netherlands

Accepted 28 June 1990

Summary

In 1985 Dutch health care authorities and health insurance companies initiated a large-scale technology assessment (TA) of liver transplantation (LTx) in The Netherlands. The 10-year experience of the existing programme in the University Hospital Groningen was investigated. Topics included were patient flow, selection policies, survival, quality-of-life, costs, need, supply of donor organs and organisational aspects. Estimation of the consequences of a non-transplantation scenario allowed for the execution of a cost-effectiveness analysis. *Results* showed clear improvement by LTx of survival and quality-of-life, though to a lesser degree than expected. Costs of the first transplantation year amounted to Dfl 180,000 (*approx* US $ 90,000). The cost-effectiveness ratio ranged from Dfl 47,000 to Dfl 133,000 per life year gained. No overt imbalance between need and donor supply existed or was expected in the near future. The *impact* of this study is related to the informational value and to the contribution to the decision-process. Even at its appearance in 1988, the final report provided health policy makers with new information. Health policy concerning LTx was considerably influenced, as a rule in agreement with the study conclusions. We conclude the Dutch case study to be an example of a useful and efficient TA.

Liver transplantation; Cost; Survival; Quality of life; Health policy; Technology assessment

Introduction

In The Netherlands orthotopic liver transplantations have been performed since 1979 in the Academisch Ziekenhuis Groningen (AZG) [1,2]. In *orthotopic* liver transplantation the diseased liver of the recipient is removed (explantation) and subsequently replaced (implantation) by a liver of approximately the same size from a donor with the same blood group. Few advocate *heterotopic* liver transplantation on some occasions. In this procedure the diseased liver is left in situ and part of a donor liver is additionally placed in the abdomen below and behind the diseased liver. In this case the donor liver is connected to the biliary and circulatory system by seperate anastomoses.

The transplantation-programme in the AZG started as an experimental clinical programme, funded by general research grants. Despite the recommendations of the Consensus Development Conference of the National Institutes of Health in 1983 and the favourable development of the programme, Dutch health care authorities were reluctant to incorporate liver transplantation (LTx) into regular reimbursement schemes [2,3]. As a result the AZG received supporting grants from the Dutch health insurance companies to continue their programme on a larger scale from 1985 onwards, one of the conditions being participation in a technology assessment (TA) by an independent research team. This technology assessment should yield objective information about costs, effects, organisational aspects, donor supply and other topics, which health care authorities regarded to be necessary for a well-founded decision [4]. The final report appeared in September 1988, describing the experience of 221 screened and 73 consecutively transplanted patients [1,5]. The results of this technology assessment were used in the subsequent decision-process. Based on the information of the final report it was decided that the restrictive regime, limiting LTx to one national centre, should be continued whilst additional research should be conducted towards information about results of LTx in the long run.

In the first part of this article the general results of the LTx-technology assessment are summarized (programme description, survival, quality of life, costs, donor supply). Also the main features of the resulting decision process will be described. In the second part two evaluative questions will be addressed as far as the state of affairs in The Netherlands in February 1990, i.e.:

1 Did the TA generate knowledge and information which might justify the research effort?
2 To what extent has the process of decision making been influenced by the information from the TA?

Finally we will discuss some critical aspects of this TA in view of future application of this policy tool.

Data and methods of the TA

Data

Medical data were provided by a clinical database describing pre-transplantation status and follow-up status of transplanted patients at regular intervals (before, and 1 week, 1 month, 1 year, 2 years, 3 years, 5 years and 10 years after transplantation; about 1100 variables per patient with complete follow-up). Medical data were complete for 76 transplantations (68 patients); from 5 patients no (sufficient) data could be obtained. Additional medical data were abstracted from medical records of chronic cirrhosis patients who were rejected for LTx to enable estimation of non-transplantation survival (131 patients) [6].

Administrative data and *socio-demographic* data were available for all referred patients, including vital status at the closing of follow-up September 1987. The following stages were distinguished: screening, waiting list, operation – 3 months, 4–12 months and every consecutive year.

Quality-of-life was measured applying computer assisted interviewing [7,8]. The interview was presented to the patient before LTx and at regular intervals afterwards, and consisted of several questionnaires representing the full spectrum of dimensions of quality-of-life [9].

Cost volume data were retrieved from the hospital administration system, which registers the majority of clinical activities (in hospital stay, physician services, lab services etc.) on a per patient per diem basis. The input volume of key specialities was determined separately on a real time basis. Other cost units counted separately concerned blood products, operating room facilities, drugs, social work and out-of-pocket patient costs. Prices (1987-level) for each cost unit (time, product or service) were derived from the financial adminstration.

Donor data were collected at Eurotransplant, an organisation which is responsible for optimal distribution and exchange of human donor organs in Western Europe.

Data on the *need* for LTx were derived from three different sources: national age- and sex specific mortality statistics, epidemiological literature and the aforementioned administrative system which included medical details for those referred patients which were regarded temporary or definitely unsuitable for LTx.

Literature on all topics was collected extensively to provide a full reference for the case results of the AZG.

Methods

Flow chart analysis and life-table techniques were used to describe the *patient flow* from the referral to post-LTx status [11]. Conventional statistics were used to compare characteristics of different patient-groups and to compare characteristics over time.

Survival analysis included Kaplan-Meier estimates of survival and Cox-regression analysis on determinants of success of the transplantation [10–14]. A new statistical technique was developed which enabled a comparison between observed

survival curves (Kaplan-Meier) and model-predicted non-transplantation survival curves [15,16].

Quality-of-life analysis included comparison of both one-dimensional and aggregate quality-of-life indicators over time [17–21]. A utility estimate was arrived at, based on multi-attribute utility theory [22,23].

Cost analysis followed methodological guidelines on economical cost-analysis [10]. Estimates of quality-of-life and costs in the hypothetical absence of the LTx-programme were derived from the waiting list period of the patient (on average 3 months).

Potential *donor supply* was extrapolated from existing trends on renal organ donation and multi-organ donation. We assumed an increasing fraction of multi-organ donations to be used for liver transplantation in the future due to relaxation of relative contra-indications and a decrease of organisational waste as a result of better preservation techniques.

Finally the *need* was estimated in three independent ways, combining a variety of sources on the suitability of potential patient groups (deaths or diseased with specific liver disease) with demographic trends.

An economical cost-effectiveness analysis was carried out separately, combining results on costs, survival and quality-of-life [24].

For the statistical analysis we used the BMDP-statistical software, release 1985 [25].

Results

Description of the LTx programme

The *treatment protocols* regarding patient selection, timing, infection prevention, surgical techniques are published elsewhere [26] and will not be discussed here. From a clinical point of view the most important identifiable innovations during the study period were the introduction of a veno-venous bypass technique and the change to a triple drug immuno-suppressive scheme [27, 28]. At the start indications were limited to adult patients with chronic cirrhosis, a clinical entity which is the final common pathway of a wide array of chronic diseases of the liver and of the biliary tract. Diseases with a high probability of recurrence (malignancies, infectious hepatitis) were excluded. Indications were gradually extended to retransplantations (1982), paediatric transplantations (1982) and, more recently, transplantations in patients with acute hepatic failure and HBsAg-positive cirrhosis, due to hepatitis B infection [5]. Gradually patients with more advanced disease were transplanted [15].

In the first 10 years of the programme, 221 referred patients satisfied criteria for further examination. About two thirds of them, 145 patients, were regarded to be suitable candidates. The difference with the number of patients transplanted at the end of data collection (Summer 1987) (73) may be explained by the large number of patients whose condition was considered to be too good for transplantion at acceptation date. Patient-referral attained a stable level in 1986, but the prediction

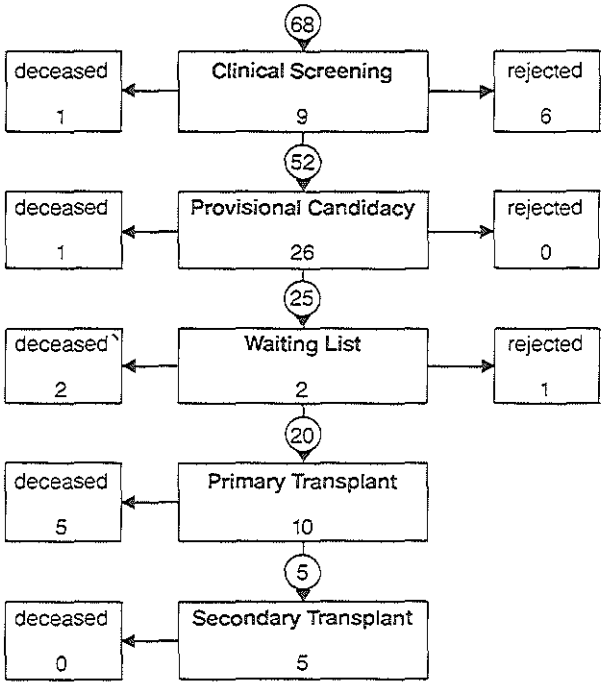


Fig. 1 Flowchart of the liver transplantation programme of the AZG 1985–1988 (01-01-1988).

was that transplantation frequency would stabilize later (about 1990).

Until 1988 the major part of accepted patients consisted of chronic cirrhosis patients (see Fig. 2). Detailed analysis of medical characteristics for primary biliary cirrhosis (PBC) and other types of chronic cirrhosis revealed some change over time: referral was more adequate (fewer rejections) and transplantation was carried out later in the course of the disease (e.g., higher Child-Pugh score, which indicates a higher degree of liver dysfunction).

International comparison was not possible for the flow-chart and the medical characteristics, but results from the European Registry for Liver Transplantation showed the Groningen patient population to be comparable to the average European patient with regard to age [5,29] and diagnosis (see Fig. 2) [29].

Only acute hepatic failure, biliary atresia (an infrequent congenital malformation of the biliary tract, characterized by interference of bile drainage, cholestasis, and, if untreated, death after weeks to months) and tumours were relative infrequent diagnoses. *Overall-survival* after LTx appeared to be 64% at one year and 53% at five year. Clearly this gives no good impression of the effect of LTx, as survival is strongly related to the nature of the pre-existent disease. In a Cox-regression model, characteristics of the recepient (age, liver function, renal function), the donor (renal

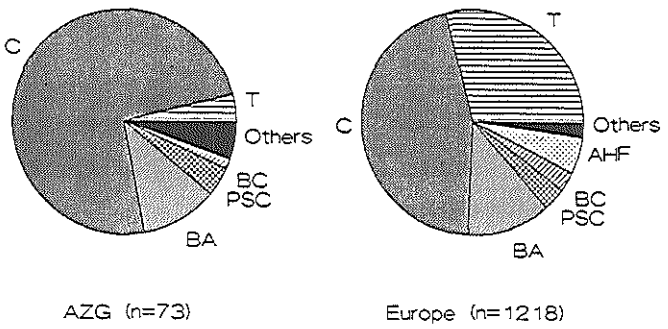


Fig. 2 Diagnosis preceding primary liver transplant; comparison of AZG and Europe (1988).T, malignant tumors; C, chronic cirrhosis, all causes; BA, biliary atresia; PSC, primary sclerosing cholan- gitis; BC, Budd Chiari syndrome; AHF, acute hepatic failure; Others, other diseases.

function) and the surgical procedure (blood loss, related to haemostatic factors) appeared to predict survival significantly.

At the start of the study no empirical information was available about the improvement of survival by LTx. The effectiveness of LTx was usually related to survival after LTx per se, assuming survival without transplantation to be negligibly low. Publications about heart transplantation (HTx) showed considerable differences between clinical predictions and empirical evidence about the prognosis in absence of HTx [30]. We therefore tried to support the clinical assumption by empirical evidence. In the absence of randomized controlled clinical trials we estimated non-transplantation survival for three diagnostic categories (primary biliary cirrhosis, chronic cirrhosis (excl. PBC) and biliary atresia) by two different methods (Fig. 3). For the two cirrhosis groups we used two published prediction models and a new model, based on the medical data of 131 non-transplanted patients in the AZG. In the case of biliary atresia we assumed the simultaneously published national experience with the Kasai-procedure, which is the therapy of first choice in the Netherlands, to provide the best control survival curve [31].

The differences in the cirrhosis-groups using the Christensen-models were smaller than expected but on the long term LTx appeared to contribute positively to survival (PBC 1 year and 5 year survival with LTx: 62 vs. 55%; without LTx: 35 vs. 26%; non-PBC-cirrhosis 1 year and 5 year survival with LTx: 60 vs. 51%; without LTx 46 vs. 21%). Patients in Child-Pugh class B and C (most deteriorated liver function) benefitted most from the transplantation, despite increased operative risk [5, 16]. The model based on non-transplanted AZG-patients yielded slightly better net effects of LTx. In the case of biliary atresia, differences appeared to be fairly large, indicating LTx (usually preceded by a Kasai procedure) to be far superior to the Kasai-procedure alone (1 year and 5 year survival 100 vs. 62% and 100 vs. 47%).

Results of *quality-of-life* analysis are shown in Table 1. Pre-transplant scores suggest major limitations on all domains of life, the most predominant feature being severe lack of energy. After transplantation all objective indicators show

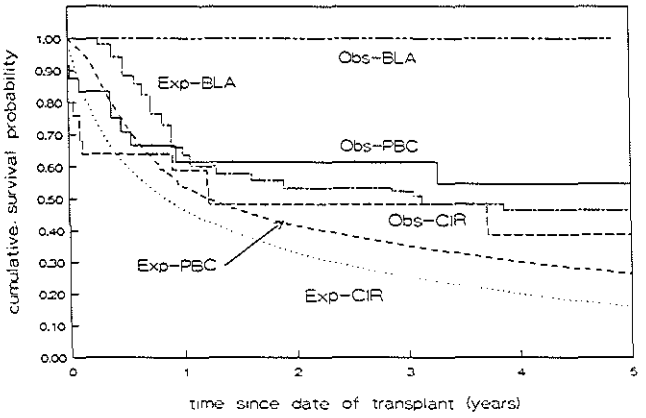


Fig. 3 Observed survival with LTx compared to expected survival without LTx in PBC, non-PBC cirrhosis and biliary atresia. Obs–PBC, observed survival with LTx in PBC; Obs–CIR, observed survival with LTx in Non-PBC cirrhosis; Obs–BLA, observed survival with LTx in biliary atresia; Exp–PBC, expected survival without LTx in PBC (cf. Christensen model); Exp–CIR, expected survival without LTx in Non-PBC cirrhosis (cf. Christensen model); Exp–BLA, expected survival without LTx in biliary atresia (cf. national Kasai-experience).

improvement, although the level of the general population is not always attained. Subjective quality-of-life attains levels which in the first year exceed the level of the general population, maybe due to euphoria of surviving the most critical surgical phase. Psychiatric morbidity was seen rarely. Until recently only few indicators of quality-of-life have been studied, usually from selected survivors after LTx. The few systematic studies available in 1988 showed conflicting results though they shared our findings of sharp improvement of subjective quality-of-life contrasting to moderate and gradual improvement of objective quality-of-life indicators [32,33]. The serious psychiatric morbidity reported by others [34–37] could not be replicated in our series. These discrepancies may be the result of a different patient selection policy or the effect of the thorough mental and social support programme in the AZG.

Costs of the programme are shown in Table 2. Clearly the first three months after transplantation are the most expensive from the perspective of cost per unit of time. However, the impact of cyclosporin (CyA) maintenance therapy to cumulative costs is impressive. For comparison we added an hypothetical CyA-conversion scheme, converting the patient to azathioprine/prednisone after one year. If we assume a survival of 20 years costs decrease with 26% (undiscounted) or 19% (discounted), respectively. Decomposition of total costs in the first year after transplantation (Dfl. 180,000, US$ 90,000) shows net costs of transplantation (25%) and hospital care (30%) (immune-suppression and medical services excluded) contributing most. Drugs account for another 12%.

Limited reference values were available in September 1988 on the costs of LTx. Reports from the U.S. showed charges for a one year surviving patient

Table 1
Quality-of-life before and from 1 year after liver transplantation

Questionnaire	Range*	Normal value	Pre	1+Yr
Karnofsky Performance Index	0–100	> 90	67	87
State-Trait Anxiety Inventory	80–20	< 37	41	32
Self-rating Depression scale	100–25	< 33	51	44
Index of Psychological Affect	1–7	> 6	4.5	6.4
Overall Life Satisfaction	1–7	> 6	3.8	6.7
Nottingham Health Profile-I	100–0	< 10		
– Mobility			25	14
– Pain			23	5
– Energy			52	6
– Sleep			39	14
– Emotional reactions			16	6
– Social isolation			14	5
Nottingham Health Profile-II**				
Present state of health causes problems for:				
– Job of work (% yes)			33	17
– Looking after the home (% yes)			67	41
– Social life (% yes)			67	13
– Home life (% yes)			50	0
– Sex life (% yes)			50	9
– Interest and hobbies (% yes)			67	22
– Holidays (% yes)			100	41
Symptoms and complaints**				
– Abdominal distension (% yes)			46	21
– Itching (% yes)			46	13
– Icterus (% yes)			61	21
– Bone pain (% yes)			38	43
– Low performance capacity (% yes)			78	17

* Theoretical range, from worst (left value) to best (right value).
** The healthy population is assumed to give the answer "no" in all instances.

ranging from $ 230,000 to $ 340,000 [38]. Additional years of survival would cost between $ 10,000 and $ 20,000 per year. These values had little meaning in the Dutch case because (a) charges are different from economical costs and (b) even applying purchase power parities, surgical costs in the U.S. are incomparably high in comparison to most other Western countries [39]. An unpublished report of the hospital administration of the AZG estimated total cost for the procedure including five year of follow-up at least being the double of our empirical results. The cost-effectiveness ratio ranged from Dfl 47,000 to Dfl 133,000 per life year gained, depending on the severity of disease at transplantation date [24]. About the same range of costs per QALY was obtained [24].

The *need* for transplantation is clearly related to many factors. Generally the epidemiology of relevant end-stage liver diseases, detection and referral, and clinical selection are the most important ones. Estimations for the Dutch case showed consistently a need of 2 to 5 per 10^6 inhabitants, given a protocol which excludes active Hepatitis B-infection, alcoholic cirrhosis and patients aged 60 years and over (see Table 3). Also the vigorous screening of candidates with a primary hepatic tumour excluded the larger part of these patients from transplantation.

Table 2
Costs (Dutch florins) of one liver transplantation patient per stage including 20 years of follow-up (*n*=76)

Transplantation stage	Undiscounted costs	5% Discounted costs
Screening/provisional candidacy	15,514	17,104
Waiting list	3,646	3,828
Operation – 3 months	132,860	132,860
4–12 months after LTx	46,318	46,318
Each consecutive year		
– with Cy-A	28,099	CF × 28,099*
– without Cy-A	12,050	CF × 12,050*
Cumulative costs (20 year survival)		
– with Cy-A	732,219 (100%)	539,695 (100%)
– without Cy-A	427,288 (74%)	345,738 (81%)

* The costs for each consecutive year are equivalent to the annual costs, corrected for the time they are incurred with the following correction factor (CF): $CF = \left(\frac{1}{1.05}\right)^{YR}$

Estimations of donor organ availability did meet these need-projections on an annual base (Table 3), but this does not preclude the occurrence of actual donor shortage (especially in the case of children).

Alternative estimates of need did exist, ranging from 4 to 20/10^6 and 20 to 250/10^6, respectively [40,41]. Only the first range is suitable for comparison as it is based on a similar epidemiological approach and on a country with about the same epidemiology of liver diseases [42]. The difference between O'Grady and our study is explained by different inclusion criteria: in the English case patients of all ages and all alcoholics were regarded potential LTx-candidates. With our inclusion criteria the estimates of need were identical for the United Kingdom and The Netherlands, given a comparable epidemiological profile of chronic liver diseases.

At the time of decision (1988) information on donor supply was only partially available. Little was known about the preventable organisational waste of organs and the potential increase of liver donation due to relaxation of suitability criteria.

Table 3
Transplantation need and donor supply in The Netherlands on an annual base

Need			Supply		
	Min	Max		Min	Max
Number provisional candidates ('84-87)	20	20	Number of MUD* (86-87)	60	60
Acute hepatic failure	+ 2	+ 5	Improved donor acquirement	× 1.00	× 1.40
Inclusion of HBsAG*	+ 2	+ 10	Change legal system	× 1.00	× 1.20
Optimal referral	× 1.0	× 1.9	Suitable liver donors	× 0.75	× 0.90
Retransplantation	× 1.1	× 1.2	Improved utilization	× 0.90	× 1.00
Crude total	26.4	79.8	Crude total	40.5	90.7
Standardized total	2/10^6	6/10^6	Standardized total	3/10^6	7/10^6

* = multi-organ donation.

Description of the decision process

In this section the judgement and decisions on LTx made by the Dutch health care authorities will be described. Before 1985 LTx was excluded from regular reimbursement. During the study-period (1985–1988) decisions of government and health insurance companies were deferred, which a.o. implied competing liver centres could not start a LTx-programme.

Before the TA started in 1985, two national advisory boards were asked to prepare advices on LTx to be released in 1988. *The National Health Insurance Board* (NHIB), which unites the insurance interests of premium parties (health care providers, employers, employees, insurance companies) set up the above-described TA to provide for the required information on costs, benefits and potential future impact. The *Health Council* (HC), which was primarily asked to report on the scientific state of the art of LTx, instituted an expert committee consisting of liver disease experts from all Dutch liver centres. A simultaneous appearance of both advices was aimed at.

Generally, the NHIB accepted all the empirical evidence of the study [43]. However, a clear decision was considered to be difficult for several reasons. First, long term results (i.e., beyond a 5-year-horizon) were still unknown, though available evidence suggested no additional risks for long term complications. Second, clear definition of selection and timing criteria seemed critical but difficult to achieve. Third, need projections were based on epidemiological data combined with rather crude estimates on the distribution of disease severity. Without data on this distribution, the size of even a clearly defined patient group seemed difficult to project. Finally, treatment policies of paediatricians with regard to biliary atresia suggested dissensus, with an unknown probability of future convergence.

Given these uncertainties and an unfavourable cost-effectiveness ratio compared to e.g., heart transplantation, the NHIB advised to postpone a definite decision for the LTx-programme as a whole. Furthermore, the TA should be continued to provide answers on the questions left. In view of the projected need and the potential capacity of the AZG-programme, the number of LTx-centres should not be extended.

The *Health Council* regarded the TA-results as only one piece of evidence [44]. The publications of the Pittsburgh centre were regarded as equivalent sources of information, though it was recognized that important views expressed in these articles sometimes lacked empirical evidence. From a scientific point of view it was concluded that LTx had left the realm of experimental care, although the expert committee acknowledged the same uncertainties as the NHIB. The critical criterion as stated in the HC-report was the survival benefit. Survival gain was regarded proven in PBC-patients and most likely in patients with other types of chronic cirrhosis and biliary atresia. Implicitly this view suggested conditional inclusion of LTx in the reimbursement schemes, as the status of LTx as an accepted and acceptable therapy could no longer be doubted. On some important clinical issues (timing of surgery, the acceptance of patients with active hepatitis B, acute hepatic failure and the preferred therapeutic strategy in biliary atresia) no statements

were made. As did the NHIB, additional research was regarded useful (e.g., on heterotopic LTx and on the need for LTx) but not mandatory for an answer on the status of the therapy. One organisational issue received much attention: the possible increase of the number of liver transplantation centres. The HC committee favoured extension of the LTx-programme to two, maybe three centres, primarily to set up a trial comparing orthotopic liver transplantation with the heterotopic procedure, which has been advocated by the Dutch liver centre in Rotterdam [45].

Both advisory reports and its conclusions – though not always concordant – were accepted by the *Minister of Health* in the summer 1989. In a letter to the parliament on the rationing of health care and more specifically on the control of new technologies as heart transplantation, in vitro fertilisation and LTx, LTx was subjected to the regimen as proposed by the NHIB. TA-research was continued and the financial and legal arrangements were kept unaltered, preventing other centres to start their own LTx-programme. The projected TA research was restricted to the questions about timing and long-term benefits. An analysis of the need for LTx was regarded unfeasible by the expected refusal of registration offices to enable analysis of death certificates and hospital records of diseased due to chronic liver diseases. Unexpectedly and still poorly understood the biliary atresia problem was regarded not important enough to justify further research.

Evaluation

Informational value

Justification of the TA of LTx in The Netherlands should be primarily related to the informational value. An important criterion in this context is the degree to which the study generated new information. New information may be defined as information which fills an information gap, challenges existing information or adds information to existing data (as e.g., in meta-analysis or cost-effectiveness analysis). Appropriate reference dates are the beginning of the study (July 1985) and the appearance of the final report (June/September 1988). In 1985 information about LTx was primarily devoted to survival after LTx, and to clinical aspects (surgery, immune-suppression). Even in 1988 comparative information on other topics was scarcely available, despite the widespread application of LTx. The greater part of the reports originated from the large LTx-centre in Pittsburgh (U.S.A.). However, many of these were of limited use either by national specifity (need, supply, costs) or by methodology (costs, quality-of-life, survival without LTx). It should be added that particularly on the subject of estimation of non-transplantation survival and quality-of-life much has been published since [16,41]. We therefore conclude the study generally satisfied the above mentioned criterion of informational value. In 1988 the informational value was not primary the result of the filling of an information gap but the result of providing independent information on several topics on which divergent views existed.

The contribution of the TA to the decision-process

The description of the policy impact of this TA should not be limited to the contribution of the final report to the explicit decision process. It should be noticed, that the research programme exerted considerable influence on the diffusion process of LTx *during* the study period. E.g. detailed analysis of medical characteristics of the patients, organisational characteristics of the programme and preliminary survival results were published as interim-reports in 1987. These provided justification of the continuation of the restrictive approach of LTx in only one centre.

At least three contributions of the final report may be distinguished. The NHIB adopted the entire contents soon after its appearance and the HC judged LTx to be acceptable and accepted therapy, based on the same evidence predominantly with regard to survival benefit. The report also convinced many in the periphery of the decision-setting of the good quality-of-life of the great majority of surviving transplanted patients. In view of the provisional acceptance of LTx one might conclude that the TA contributed to the decision that LTx was an *accepted form of therapy*.

Second, the analysis of the need for LTx left little room for a second LTx-centre in The Netherlands, from a health care point of view, given the fact that even an excentrically located unit has good accessibility due to the small size of the country and excellent transport facilities. The continuation of *restrictive legislation* on the *number of centers* was clearly based on this information.

Third, the final report provided some suggestions on improvement of the referral and selection process, which clearly influenced the *practice* of LTx in the AZG.

Discussion

An appropriate judgement of the results and impact of the TA of the Dutch LTx-programme should consider some important aspects of TA beyond the informational value as such. These aspects are equally decisive for the viability of future application of this policy tool. Five of these aspects will be discussed.

In the section above it was concluded that the information on survival and – maybe – quality-of-life contributed to the provisional *acceptance*, thus proving the informational value. But the *provisional* character deserves attention. The NHIB suggested to postpone decisions until uncertainties had been clarified. The particular uncertainties involved (selection, need, long-term effects) will exist at any time with any recent technology. The requirement of certainty may be regarded as a *process*-criterion, not a mere outcome-criterion, which neglects the inherent value of the technology sofar demonstrated, as in the case of LTx in children with biliary atresia. The emphasis on process-criteria is related to the use of TA as a process management tool (usually a 'freezing tool') rather than as a decision supporting tool. This use illustrates a yet immature decision process.

From a health insurance perspective another undesirable aspect of the decision-

process is the lack of information on cost-effectiveness of other health care programmes. LTx was apparently less cost-effective as compared to heart transplantation (HTx) but due comparison would include interventions for other chronic diseases as well, notably malignancies. This implies a balanced cost-effectiveness analysis research programme on a larger scale, and a continuous decision process in the context of an increased awareness of the necessity of rational choices in health care [32]. This should be preferred to an accidental comparison of two health care programmes (HTx and LTx) which mainly share public appeal as banners of medical progress and golden standards of increase of medical expenses.

Next, the societal or public health perspective usually ignores the scientific interest of a health care programme. Regulations which aim at optimization of capacity usually conflict with maintenance of professional autonomy and the freedom of scientific research. The subtle differences between the HC-advice and the NHIB-advice clearly originate from this difference. In our opinion *patients* interests should guide the decision-maker, which implies a tendency to centralization of complex procedures of the number in centers with a capacity guaranteeing scientific and economic effeciency.

The uncertainty of the *need estimates* may be criticized. In our view this is not primarily a specific feature of TA methodology but a characteristic of descriptive population epidemiology. Population data on the incidence of many adult diseases are scarce, a result of the cost-intensiveness of this type of epidemiological research. Even more scarcely available are population data on stage or severity of disease. Mortality data or clinical admission data sometimes may be a good proxy for data on severely ill patients, but in many countries reliable nation-wide registrations do not exist or data are not available for case-by-case exploration. The development of a feasible approach of this recurrent problem is a major challenge for epidemiologists, notwithstanding the fact that one aspect of uncertainty will remain in the context of TA. Projections of future need will always be prone to changes due to developments in the technology involved and its alternatives, i.e., in this case transplantation and conventional medical therapy.

Finally, the deployment of resources and the inconvenience involved by a large scale TA should be explicitly valued as a *disutility*. No standards exist about the appropriate maximum of resources devoted to a single TA. From an economical perspective these may be related to the *economical decision space*, given a time horizon of 5 to 10 years. In our case the lower bound of the decision space was determined by a liver programme without LTx and with conventional therapy. The upper bound of the decision space was the execution of about 40 LTx each year, follow-up included. Using a short 5 year time horizon, the TA-costs made up only 3–4% of the differential costs of these two options. From the clinician's perspective the intrusion of the TA on the clinical programme was regarded with suspicion at the start. It was valued positively in the end, probably because inconveniences were balanced by scientific spin-off and an improved position on the Dutch market.

The TA of the Dutch LTx-programme taught us many lessons, one of them being that high impact technologies may be fruitfully analyzed this way.

Acknowledgements

The authors wish to thank the liver transplantation team of the Academisch Ziekenhuis Groningen (coordinator Professor M.J.H. Slooff M.D., Ph.D.) for their on-going fruitful cooperation. This study was financed by the Dutch National Health Insurance Board (Ziekenfondsraad)

References

1 Gips, C.H., Krom, R.A.F., Houthoff, H.J., Schuur, K.H., The process of selection and referral of candidates for liver transplantation (in Dutch), Nederlands Tijdschrift voor Geneeskunde, 125 (1981) 875–878.
2 Scharschmidt, B.F., Human liver transplantation: analysis of data on 540 patients from four centers, Hepatology, 4 (1984) 95S–101S.
3 National Institutes of Health consensus development conference statement: Liver transplantation, June 20–23, 1983, Hepatology 4 (1984) 107S–110S.
4 Department of Welfare, Health and Cultural Affairs. The limits of health care. Policy statement on the advices of the National Health Insurance Board (Ziekenfondsraad), the National Board on Public Health (Nationale Raad voor de Volksgezondheid) and the Health Council (Gezondheidsraad). Parliamental papers, 1987–1988, 20620, numbers 1 and 2.
5 Bonsel, G.J., Veer, F. Van 't, Bot, M.L., Costs and effects of liver transplantation, volume 4: Prognosis of liver transplantation, the Groningen programme 1977–1987 (in Dutch), Rotterdam, Institute of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands, 1988.
6 Kalbfleisch, J.D., Prentice, R.L., The statistical analysis of failure time data, New York, John Wiley & Sons, 1980.
7 Erdman, H.P., Klein, M.H., Griest J.H., Direct patient computer interviewing, Journal of Consultations in CliniricalPsychology, 53 (1985) 760–773.
8 Skinner, H.A., Allen, B.A., McIntosh, M.C., Palmer, W.H., Lifestyle assessment: applying microcomputers in family practice, British Medical Journal, 290 (1985) 212–216.
9 De Haes, J.C.J.M. and Van Knippenberg, F.C.E., The quality of life of cancer patients: a review of the literature, Social Science andMedicine, 20 (1985) 809–817.
10 Drummond, M.F., Stoddart, G.L., Torrance, G.W., Methods for the economical evaluation of health care programmes, Oxford, Oxford University Press, 1987.
11 Evans, R.W. and Maier, A.M., Outome of patients referred for cardiac transplantations, Journal of the American College of Cardology, 8 (1986) 1312–1317.
12 Kaplan, E.L. and Meier, P., Nonparametric estimation from incomplete observation, Journal of the American Statistical Association, 53 (1985) 457–81.
13 Harrington, D.P., Fleming, T.R., A class of rank test procedures for censored survival data, Biometrika, 69 (1982) 553–566.
14 Turnbull, B.W., Brown, B.W., Hu, M., Survivorship analysis of heart transplant data, Journal of the American Statistical Association, 69 (1974) 74–79.
15 Christensen, E., Multivariate survival analysis using Cox's regression model, Hepatology, 7 (1987) 1346–1358.
16 Bonsel, G.J., Klompmaker, I.J., van't Veer, F., Habbema, J.D.F., Slooff, M.J.H., Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis, Lancet, i (1990) 493–497.
17 Grieco, A. and Long, C.J., Investigation of the Karnofsky performance status as a measure of quality of life, Health Psychol 3 (1984) 129–142.
18 Spielberger, C.D., Gorsuch, R.L., Lushene, R.E., STAI manual for the state-trait anxiety inventory, Palo Alto, California, Consulting Psychologists Press Inc., 1970.
19 Zung, W.W.K. and Durham, N.C., A self-rating depression scale, Archives of General Psychiatry, 13 (1965) 63–70.
20 Hunt, S.M., McEwen, J., McKenna, S.P., Measuring health status, Croom Helm, London, 1986.
21 Campbell, A., Converse, P.E., Rodgers, W.L., The quality of American life, Russel Sage Foundation,

New York, 1976.
22 Torrance, G.W., Measurement of health utilities for economic appraisal, a review, Journal of Health Economics, 5 (1986) 1–30.
23 Boyle, M.H., Torrance, G.W., Sinclair, J.C., et al., Economic evaluation of neonatal intensive care of very-low-birth-weight infants, New England Journal of Medicine, 308 (1983) 1330–1336.
24 Bonsel, G.J., Klompmaker, I.J., Bot, M.L. et al. Cost-effectiveness analysis of the Dutch liver transplantation programme, Transplantation Proceedings, 22 (1990) 117–120.
25 Dixon, W.J. (ed), BMDP Statistical Software, 1983 printing with additions, Los Angeles, University of California Press, 1983.
26 Van der Putten, A.B.M.M., Bijleveld, C.M.A., Slooff, M.J.H. et al., Selection criteria and decisions in 375 patients with liver disease considered for liver transplantation during 1977–1985, Liver, 7 (1987) 84–90.
27 Slooff, M.J.H., Barns, J.L., Sluiter, W.J., Klompmaker, I.J., Hesselink, E.J., Verwer, R., A modified cannulation technique for veno-venous bypass during orthotopic liver transplantation, Transplantation Proceedings, 21 (1989) 2328–2329.
28 Klompmaker, I.J., Haagsma E.B., Gouw, A.S.H., Verwer, R., Slooff, M.J.H., Azathioprine and prednisolone immunosuppression versus maintenance triple therapy including cyclosporin for orthotopic liver transplantation, Transplantation, 48 (1989) 814–818.
29 Bismuth, H., Ericzon, B.G., Rolles, K., Castaing, D., Otte, J.B., et al., Hepatic transplantation in Europe, First report of the European liver transplant registry, Lancet, i (1987) 674–676.
30 Buxton, M., Acheson, R., Caine, N., Gibson, S., O'Brien, B., Costs and benefits of the heart transplant programmes at Harefield and Papworth hospitals,: Final Report, Department of Economics, Brunel University and Department of Community Medicine, University of Cambridge, 1985.
31 Houwen, R.J., Zwierstra, R.P., Severijnen, R.S.V.M., Bouquet, J., Madern, G., Vos, A. et al., Extrahepatic biliary atresia in The Netherlands. The long-term prognosis after surgery, Archives of Disease in Childhood, 64 (1989) 214–218.
32 Leaf, A., Cost effectiveness as a criterion for Medicare coverage, New England Journal of Medicine, 321 (1989) 898.
33 Williams, J.W., Vera, S., Evans, L.S., Socioeconomic aspects of hepatic transplantation, American Journal of Gastroenterology 82(1987) 1115–1173.
34 Penn, I., Psychiatric aspects of hepatic transplants. In: Howells J.G., Modern perspectives in psychiatry, New York, Brunner & Mazel, 1976.
35 House, R., Dubovsky, S.L., Penn, I., Psychiatric aspects of hepatic transplantation, Transplantation, 36 (1983) 146–150.
36 Tarter, R.E., Thiel, D.H. van, Hegedus, A.M., Schade, R.R., Gavaler, J.S., Starzl, T.E., Neuropsychiatric status after liver transplantation, Journal of Laboratory and Clinical Medicine,776–782.
37 Surman, O.S., Dienstag, J.L., Cosimi, A.B., Chauncey, S., Russell, P.S., Psychosomatic aspects of liver transplantation, Psychotherapy and Psychosomatics, 48 (1987) 26–31.
38 Krueger, H., Economic analysis of solid organ transplantation: a review for policy makers, Health Policy, 13 (1989) 1–17.
39 Iglehart, J.K., Japan's medical care system (second of two parts), New England Journal of Medicine, 319 (1988) 1166–1173.
40 O'Grady, J.G., Williams, R., An appraisal of liver transplantation in Great Britain, International Journal of TechnicalAssesment in Health Care, 2 (1986) 465–469.
41 Starzl, T.E., Demetris, A.J., Van Thiel, D., Liver transplantation (first of two parts), New England Journal of Medicine, 321 (1989) 1014–1022.
42 Massé, L., Juillan, J.M., Chisloup, A., Trends in mortality from cirrhosis of the liver 1950–1971, World Health Statistics Report, 29 (1976) 40–67.
43 National Health Insurance Board, Liver transplantations, Advice on the reimbursement of liver transplantations, NHIB-paper number 429, Amstelveen, 1988.
44 Health Council, Liver transplantation, Advice to the Minister of Welfare, Health and Cultural Affairs, The Hague, 1989.
45 Terpstra, O.T., Schalm, S.W., Weimar, W., Willemse, P.J.A., Baumgartner, D., Groenland, T.H.N. et al., Auxiliary partial liver transplantation for end-stage chronic liver disease, New England Journal of Medicine, 319 (1988) 1507–1511.

Cost-Effectiveness Analysis of the Dutch liver Transplantation Programme

G.J. Bonsel
I.J. Klompmaker
M.L. Essink-Bot
J.D.F. Habbema
M.J.H. Slooff

Transplantation Proceedings, Vol 22, No 4 (1990) 1481-1484
Published with permission of the editor

Cost-Effectiveness Analysis of the Dutch Liver Transplantation Programme

G.J. Bonsel, I.J. Klompmaker, M.L. Essink-Bot, J.D.F. Habbema, and M.J.H. Slooff

LIVER transplantation (LTx), though never assessed in a controlled clinical trial, is performed in growing numbers in centres all over the world.[1] As the costs of the procedure including follow-up treatment are impressive, health insurance companies have been reluctant to incorporate LTx into their insurance scheme.[2-4]

In The Netherlands, orthotopic liver transplantations have been performed since 1979 in the Academisch Ziekenhuis Groningen.[5]

In view of the unknown costs and effects of the liver transplantation programme in The Netherlands, the Dutch government and the National Health Insurance Board (Ziekenfondsraad) agreed in provisionally subsidizing the AZG-programme. In 1985 the AZG was provided with a 3-year grant, on the condition that the hospital would participate in a medical technology assessment by an independent research team. This study should aim at an empirical description of the costs, effects, organisational impact and other aspects of the AZG-programme to support a decision about incorporation in the Dutch health insurance scheme (Table 1).

Legal and ethical issues and the concluding health policy recommendations were dealt with by two national advisory boards. The final report of the study appeared (in Dutch) in September 1988, describing the experience of 221 screened patients and 81 transplantations (including 8 retransplantations) between 1978 and 1987.[6] This article presents a cost-effectiveness analysis based on the data of this study.

PATIENTS AND METHODS

Treatment Protocol

Treatment protocols regarding patient selection, timing, infection prevention, surgical technique, and immunosuppression are described elsewhere.[7] Changes in the study period included the introduction of a modified veno-venous bypass technique (1982)[8] and the change to a triple drug immunosuppressive scheme (1985).[9] Cost calculations were based on a treatment protocol including cyclosporin A.

Table 1. Topics of the Dutch Orthotopic Liver Transplantation Assessment

Quantitative Analysis	Additional Topics
Survival with and without LTx	Patient criteria
Quality of life with and without LTx	Process of referral, screening, timing
Costs with and without LTx	Organisational aspects
Cost-effectiveness of LTx	Changes over time
Need for LTx and supply of donor livers	

LTx = liver transplantation.

Medical Data

A medical database (max 1100 variables per patient) described pretransplantation status and follow-up status of transplanted patients at regular intervals. Medical data were complete for 76 transplantations.

Administrative and socio-demographic data were available, including the vital status at the closing of follow-up (autumn 1987). The following stages were distinguished: screening; waiting list; operation to 3 months; 4 to 12 months, second year, third year, etc.).

Additional data were abstracted from medical records of chronic cirrhosis patients who were rejected for LTx to enable the modelling of nontransplantation survival by means of Cox-regression analysis (n = 131).[10]

Quality of Life Data

Quality of life and health status were measured applying computer-assisted interviewing. The time schedule included an interview before LTx (waiting list) and at regular intervals after LTx (3 months, 1-year, 2-year, and consecutive years).[11,12] The interview consisted of several questionnaires representing the full spectrum of objective and subjective measures of quality of life.[13] Children were excluded (age limit 16 years).

Cost Data

Cost volume data were abstracted from the hospital administration system, which registrates the majority of clinical activities (admissions, physician services, laboratory services, etc.) on a per patient per diem base.

The input volume of key specialties (hepatology, paediatrics, surgery, anaesthesiology) was determined separately, based on real time devoted by the medical staff to various tasks (clinical, outpatients, overhead). Other cost units counted were blood products, operating room facilities, donor procurement, drugs, social work, and out-of-pocket patient costs. Capital outlays were estimated on a annual programme base by data provided by the financial administration of the hospital. Prices (1987-level) for

From the Institute for Medical Technology Assessment (G.J.B.) and Department of Public Health and Social Medicine (G.J.B., M.L.B., and J.D.F.H.), Medical Faculty, Erasmus University, Rotterdam, The Netherlands and the Liver Transplant Group AZG (I.J.K. and M.J.H.S.), Department of Internal Medicine (I.J.K.), Department of Surgery (M.J.H.S.), Academisch Ziekenhuis Groningen, The Netherlands.

This study was financed by the Ziekenfondsraad (the Dutch National Health Insurance Board).

Address reprint requests to G.J. Bonsel, MD, Institute for Medical Technology Assessment, Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands.

0041-1345/90/$3.00/+0

each cost unit (time, product, or service) were derived from the financial administration of the hospital.

Methods

Conventional descriptive statistics were computed with the BMDP-software package.[14] All values presented are averages. The data justified a time horizon of 5-year follow-up after transplantation.

Cost-effectiveness analysis was carried out according to the guidelines set out by Drummond et al.[15] One of the requirements is the estimation of key parameters (survival, quality of life, costs) in the absence of a liver transplantation programme.

Observed survival curves after grafting were calculated using the Kaplan–Meier estimate.[22] Two models published by Christensen were used to predict nontransplantation survival of PBC and non-PBC chronic cirrhosis patients separately.[16–18] The combined actual transplantation survival was compared with the combined expected nontransplantation survival.[19–21] For a check of validity a third model was developed based on Cox-regression analysis of the survival experience of those patients with PBC and non-PBC chronic cirrhosis, who were not accepted for transplantation.[20]

Quality of life summary data were calculated following the guidelines of each questionnaire involved.[25–27] An average utility index for pre- and posttransplantation health status was calculated based on health scenario valuations by a sample of the general public.[28] For the cost-effectiveness analysis it was assumed quality of life in absence of liver transplantation could be represented by pretransplantation waiting list values.

Cost were calculated both as undiscounted costs per stage (eg, screening period, waiting list period) and as costs per patient transplanted. Stage costs enable, for example, the comparison of the impact of various stages. Costs per transplantation are useful for planning purposes and necessary for calculation of a cost-effectiveness ratio.[15] In this case, costs incurred by patients finally not transplanted are added to the transplantation costs, weighted for their frequency relative to one transplanted patient. The transplantation date was chosen as the point in time for which the so-called "present value" of costs per transplanted patient was calculated.[15] Thus a negative discount rate was assigned to costs per transplanted patient before transplantation (screening, waiting list). An annual discount rate of 5% was chosen, following national guidelines.

For the cost-effectiveness analysis pretransplantation costs were restricted to those costs pertinent to transplantation (eg, excluding conventional treatment costs). Costs in absence of transplantation were estimated by multiplication of the predicted nontransplantation survival time (in days) with average per diem costs for conventional treatment during the waiting list period.

RESULTS

On December 1, 1987, a total of 221 patients were accepted for screening. About two-thirds were provisionally accepted for transplantation; 73 patients were transplanted, 8 of them twice. Primary biliary cirrhosis ($n = 25$) and other chronic cirrhoses ($n = 29$) together accounted for 74% of the primary transplantations.

Survival analysis of cirrhosis patients showed a significant improvement due to transplantation (log rank $P < 0.01$; see Fig 1). A positive net benefit of average years survived emerges after about 1 year, as the surgical risks

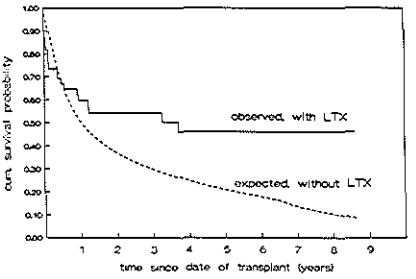


Fig 1. Survival with and without transplantation of all chronic cirrhosis patients ($n = 54$, 1978-1987, AZG). Model-predicted nontransplantation survival. *x*-axis: time since date of transplant (years); *y*-axis: cumulative survival probability.

outweigh the nontransplantation mortality in the first half year. Subgroup analysis showed no benefit for patients in Child–Pugh class A and a clear benefit for Child–Pugh class B and C patients (see Fig 2).

Pretransplantation quality of life data showed moderately impaired functional performance, severe anxiety and depression and many symptoms of chronic liver disease, especially fatigue and lack of energy. Three months after transplantation, quality of life of the survivors showed an impressive improvement; further improvement to normal or slightly affected values is seen after 1 year (see Table 2 for a selection of outcome variables). Two major psychiatric episodes were encountered (1 suicide).

Cost analysis of the various transplantation stages (nondiscounted) showed the first 3 months after transplantation accounting for about Hfl 133,000, assuming a 15% retransplantation rate (Table 3; 1 Hfl ~ 0.5 US$).

The surgical costs (donor procurement, hepatectomy,

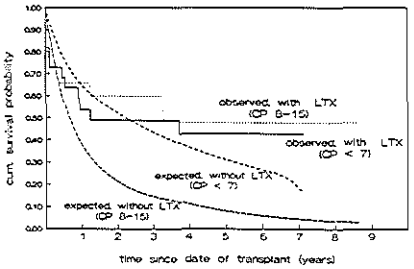


Fig 2. Survival with and without transplantation of cirrhosis patients according to Child–Pugh class (A = 22, B and C = 29, 1978-1987, AZG). *x*-axis: time since date of transplant (years); *y*-axis: cumulative survival probability.

Table 2. Quality of Life Before and After Liver Transplantation (n = 23)

Questionnaire	Range	Normal Value	Pre-LTx	3 Months	1+ Year
Karnofsky Performance Index	0-100	≥90	87	74	87
State-Trait Anxiety Inventory (anxiety)	80-20	≤37	41	33	32
Symptoms and complaints	9-0	0	4	2	2
Nottingham Health Profile-I	100-0	<10			
mobility			25	17	14
pain			23	8	5
energy			52	16	6
sleep			39	5	14
emotional reactions			16	3	6
social isolation			14	1	5
Index of well-being	2.1-14.7	>12	8.7	12.9	13.8
Utility Index	0.0-1.0	1.0	0.55	0.80	0.85

Range = theoretical range, from worst (left value) to best (right value).

implantation and additional perioperative procedures) accounted for Hfl 39,000. Follow-up costs from 1 year after LTx onwards accounted for Hfl 28,000, to which cyclosporin contributed Hfl 16,000 (about 60%). Total costs per transplantation including 5-year follow-up amounted to Hfl 250,000 (see Table 3). Costs without transplantation were estimated to Hfl 60,000, yielding net costs of Hfl 190,000 per transplantation. In view of patient 25% variability margins seemed justified, resulting in Hfl 142,500 to Hfl 237,500.

As reliable data about survival probabilities after 5-year survival were lacking, we assumed a range of 2.5 to 20 additional life years gained for those patients surviving 5 years.[29] After discounting and multiplication with the proportion surviving 5 years, 1 to 5 additional life years (at the expense of Hfl 28,000 to Hfl 140,000) are gained per transplantation, which should be added to 1 life year gained in the first 5 years (per transplantation, discounted, child B and C).

Combining the unfavourable assumptions and the favourable assumptions yielded a cost-effectiveness ratio ranging from (237,500 + 28,000)/(1 + 1) to (142,500 + 140,000)/(1 + 5) which is Hfl 133,000 to 47,000 per life year gained. When quality of life adjusting was applied results were similar: Hfl 133,000 to 51,000 per Quality Adjusted Life Year gained.

DISCUSSION

This article presents the results of the first cost-effectiveness analysis of liver transplantation. As no controlled clinical trial has been performed, we could not dispose of an unbiased estimate of control survival, quality of life, and costs. However, restricting our analysis to adult patients with chronic nonalcoholic cirrhosis we were able to approximate the nontransplantation case.

As we are aware of methodological flaws inherent to each of the approximations, validity was checked if possible. We compared the Christensen models with the AZG general cirrhosis model (rejected patients) and, for PBC patients, with the recently published Mayo-model.[20,30] As for PBC patients, the two alternative models predicted nearly identical nontransplantation survival curves, which were slightly more favourable than estimates from Christensen's model. Comparison of Christensen's model for non-PBC cirrhosis with the AZG-model showed a small difference in the same direction. These differences are

Table 3. Costs at 1987 Prices per Transplantation Stage and per Transplantation (n = 76, 1978-1987, AZG)

Transplantation Stage	Undis costs/ Pt/Stage	Disc Costs/ Pt/Stage	Patients/ Stage	Costs/ Stage
Screening	15,514	17,104	1.30	22,235
Waiting list	3,646	3,924	1.20	4,316
Operation–3 months	132,860	132,860	1.00	132,860
4-12 months after LTx	46,318	46,318	0.75	34,739
2nd year after LTx	28,099	26,761	0.60	16,057
3rd year after LTx	28,099	25,487	0.55	14,018
4th year after LTx	28,099	24,273	0.55	13,350
5th year after LTx	28,099	23,117	0.55	12,714
Total costs per transplantation at transplantation date ~ Hfl				250,000

Undis costs/pt/stage = costs per patient per stage, not discounted.
Disc costs/pt/stage = costs per patient per stage, 5% discounted relative to transplantation date.
Prices in Dutch florins; 1 Hfl ~ 0.5 US $.

maybe caused by improved conservative treatment since the time of Christensen's study.[21]

Our assumption of control quality of life may be overestimated, but while waiting list patients are generally in poor health, potential long-term survivors without transplantation might experience temporary improvement, as we observed in some patients leaving the waiting list for this reason. The hospitalization rate underlying our nontransplantation cost assumptions may be an underestimate in as far in a nontransplantation scenario patients might be subject to heroic treatment efforts, for example, for gastrointestinal bleeding. However from the past we know that many cirrhotic patients die suddenly, without prior hospitalization.

The transplantation survival curves show early mortality to be an important limiting factor for survival gain. Optimal timing combined with a decrease in perioperative mortality can further improve cost-effectiveness.

Application of our results requires outcome and cost estimates to be comparable with other reports. Though we did not find major differences with regard to the outcome, a striking difference occurs between Dutch and American cost figures.[31] This is probably caused by the fact that the frequently used financial costs (based on a.o. charges) and the real economical costs used in this study are not interchangeable concepts.[15] Also, the American price level of medical services is relatively high.[32]

Cost-effectiveness analysis is currently recognized as an additional tool for societal decisions about the choice or rejection of particular technologies.[33] Our study shows liver transplantation to have an acceptable ratio, with a lower range comparable to, for example, the cost-effectiveness ratio of heart transplantation, which in a recent Dutch study was estimated to be Hfl 50.000 per life year gained.[34]

REFERENCES

1. Starzl TE, Demetris AJ, Van Thiel D: N Engl J Med 321:1014, 1989
2. Cuthbert JA: Am J Med Sciences 291:286, 1986
3. Chu F, Cotter D, Hamilton MAT: First-year report on adult liver transplantation operating costs, medicare payments and utilization rates. Washington DC, Medical Technology and Practice Patterns Institute, Inc., 1988
4. Starzl TE, Shapiro R, Teperman L: Transplant Proc 21:3432, 1989
5. Krom RAF, Gips CH, Newton D, et al: Transplant Proc 15:1276, 1983
6. Bonsel GJ, Habbema JDF, Bot ML, et al: Nederl T Geneeskd 133:1406, 1989
7. Putten ABMM van der, Bijleveld CMA, Slooff MJH, et al: Liver 7:84, 1987
8. Slooff MJH, Bams JL, Sluiter WJ, et al: Transplant Proc 21:2328, 1989
9. Klompmaker IJ, Haagsma EB, Gouw ASH, et al: Transplantation 48:814, 1989
10. Kalbfleisch JD, Prentice RL: The Statistical Analysis of Failure Time Data. New York: Wiley, 1980
11. Erdman HP, Klein MH, Griest JH: J Consult Clin Psychol 53:760, 1985
12. Skinner HA, Allen BA, McIntosh MC, et al: Br Med J 290:212, 1985
13. Haes JCJM de, Knippenberg FCE van: Soc Sci Med 20:809, 1985
14. Dixon WJ (ed): BMDP Statistical Software, 1983 printing with additions. Los Angeles: University of California Press, 1983
15. Drummond MF, Stoddart GL, Torrance GW: Methods for the Economical Evaluation of Health Care Programmes. Oxford: Oxford University Press, 1987
16. Christensen E, Neuberger J, Crowe J, et al: Gastroenterology 89:156, 1985
17. Christensen E: Hepatology 7:1346, 1987
18. Christensen E: Dan Med Bull 35:167, 1988
19. Harrington DP, Fleming TR: Biometrika 69:553, 1982
20. Bonsel GJ, Veer F van 't, Bot ML: The Costs and Effects of Liver Transplantation, Vol 4: Prognosis. Rotterdam: Institute of Public Health, Erasmus Universiteit, 1988
21. Markus BH, Dickson ER, Grambsch PM, et al: N Engl J Med 320:1709, 1989
22. Kaplan EL, Meier P: J Am Stat Assoc 53:457, 1985
23. Grieco A, Long CJ: Health Psychol 3:129, 1984
24. Spielberger CD, Gorsuch RL, Lushene RE: STAI Manual for the State–trait Anxiety Inventory. Palo Alto: Consulting Psychologists Press, 1970
25. Zung WWK, Durham NC: Archiv Gen Psych 13:63, 1965
26. Hunt SM, McEwen J, McKenna SP: Measuring Health Status. London: Croom Helm, 1986
27. Campbell A, Converse PE, Rodgers WL: The Quality of American Life. New York: Russell Sage Foundation, 1976
28. Boyle MH, Torrance GW, Sinclair JC: N Engl J Med 308:1330, 1983
29. Starzl TE, Demetris AJ, Van Thiel D: N Engl J Med 321:1092, 1989
30. Dickson ER, Grambsch PM, Fleming TR, et al: Hepatology 10:1, 1989
31. Bismuth H, Ericzon BG, Rolles K, et al: Lancet ii:674, 1987
32. Iglehart JK: N Engl J Med 319:1166, 1988
33. Leaf A: New Engl J Med 321:898, 1989
34. Charro FT, Bonsel GJ, Hout BA van: The Costs and Effects of Heart Transplantation. Rotterdam: Faculty of Law, Erasmus Universiteit, 1988

Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis

Gouke J. Bonsel
Ids J. Klompmaker
Fred van 't Veer
J. Dik F. Habbema
Maarten J.H. Slooff

The Lancet (1990); 335: 493-97
Published with permission of the editor

Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis

GOUKE J. BONSEL IDS J. KLOMPMAKER FRED VAN 'T VEER
J. DIK F. HABBEMA MAARTEN J. H. SLOOFF

To examine the effectiveness of liver transplantation (LTx) for the treatment of primary biliary cirrhosis (PBC) the actual survival of 30 PBC patients who received liver grafts was compared with predictions of what survival would have been without transplantation. Three models, based on Cox' regression analysis, were used. Two models were derived from survival of PBC patients in drug trials and the third from cirrhotic patients who did not receive transplants. Observed and expected survival were compared for a follow-up time of 7 years. After 1 year the difference in favour of LTx was small, but after 5 years survival with LTx exceeded all predicted survival probabilities without LTx. After 3 years every year of follow-up added about 0·3 years to expected survival gain per transplanted patient, resulting in 1·5 to 2·3 life-years gained at 7 years' follow-up, depending on the model used. The benefit was greatest for patients in Child-Pugh classes B and C. The consistency between the three models in their predictions supports the validity of the use of predictive models in the indirect assessment of LTx.

Lancet 1990; 335: 493–97.

Introduction

Although never assessed in a randomised clinical trial, liver transplantation (LTx) is supposed to be effective in patients with end-stage liver disease[1–3] and may now be regarded as accepted therapy for non-alcoholic liver diseases, including primary biliary cirrhosis (PBC).[4,5]

Without information on cost-effectiveness of the procedure, national authorities tend to control the application of LTx, by financial or other restrictions, as occurs in the Netherlands, for example.

In the Academisch Ziekenhuis Groningen (AZG) orthotopic liver transplantations have been done since 1979,[6,7] and these are subsidised by the Dutch Government and the National Health Insurance Board (Ziekenfondsraad). From 1985 to 1988 liver-transplantations underwent a comprehensive medical technology assessment by an independent research group, to estimate empirically the costs and the effects of the procedure and the future need for LTx.[8,9]

This article presents our assessment of the value of LTx in prolonging survival of patients with primary biliary cirrhosis. In Europe this disease accounts for 20% of liver transplants, in AZG for about 30%.[10] For non-transplant patients three models were used to predict patient-specific survival.

Patients and methods

Patient population

Between the start of the LTx-programme in 1978 and March, 1989, 30 patients (28 women, 2 men) with PBC received liver transplants. 12 (40%) patients were in Child-Pugh class A, 11 (37%) in class B, and 7 (23%) in class C.[11,12] The mean age at transplantation was 49·6 years (SD 4·7).

Treatment protocols regarding patient selection, timing, infection prevention, surgical technique, and immunosuppression have been described elsewhere.[13] Changes in the study period included the introduction of a modified veno-venous bypass technique (1982)[14] and the change to a triple-drug immunosuppressive scheme including cyclosporin A (1985).[15]

For the medical technology assessment in 1988 we set up a comprehensive database containing clinical, biochemical, histological, and other patient data, including surgical procedure and donor. Pretransplantation data accounted for about 100 different variables and included data necessary for the application of the three prognostic models. Pretransplantation status was defined as the state of the patient at the time of transplantation.

Median follow-up after transplantation was 2·5 years, maximum 9·9 years. Final follow-up date was March 1, 1989.

Prognostic models

The Christensen model was based on data of a multicentre trial examining the effectiveness of azathioprine.[16] The AZG model was developed as part of our study and was based on non-transplanted patients with either PBC or non-alcoholic cirrhosis.[17] The third, the Mayo-model,[18] was based on a study similar to Christensen's. All three statistical models were based on Cox' proportional hazards model. Each of the models can be used to calculate a prognostic index (PI) for any patient. The higher is the value of the PI, the shorter the expected survival. Combined with the so-called underlying cumulative hazard function, the probability that a patient with a given set of values of the prognostic variables at t_0 will survive to a given time t can be estimated.

In the Christensen-model six prognostic variables are combined to give a prognostic index (PI_c) as follows:[16]

$PI_c = 2{\cdot}52$ ($\log_{10}$ serum bilirubin, μmol/l) + 0·0069 exp ([age, yr, −20]/10) − 0·05 (serum albumin, g/l) + 0·88 (cirrhosis, 1/0) + 0·68 (histological cholestasis, 1/0) + 0·52 (no azathioprine = 1, azathioprine = 0).

The AZG-model started with all 131 patients with non-alcoholic cirrhosis and primary biliary cirrhosis who, for various reasons, were excluded from the LTx programme or who were still under investigation. Nine variables were significantly related to prognosis and were selected to yield a prognostic index[17] as follows:

$PI_a = 0{\cdot}0065$ (serum bilirubin, μmol/l) + 0·0605 (age, yr) − 0·0517 (serum albumin, g/l) + 1·1827 (HBsAg) + 2·0849 (neurological complications, 1/0) + 1·2804 (varices 1/0) + 0·1866 (Quick-time prolongation, s) + 0·9183 (ascites, 1/0) + 0·7468 (clinical icterus, 1/0).

In the Mayo-model five independent variables were used for the prognostic index (PI_m) as follows:[15]

$PI_m = 0{\cdot}871$ (log$_e$ serum bilirubin, mg/dl) + 0·039 (age, yr) − 2·53 (log$_e$ serum albumin, g/dl) + 2·38 (Quick-time absolute, s) + 0·859 oedema score, 1/0·5/0),

where oedema score is 0 if no diuretic therapy was needed for oedema or no oedema was present, 0·5 if oedema resolved with diuretic therapy or if no diuretic therapy was prescribed for oedema, and 1 if oedema was present despite diuretic therapy.

Model-specific survivor cumulative hazard functions were estimated with polynomial or Weibull functions. 7 years was the maximum time for which the predictions were valid.

Statistical methods

Observed survival time was the interval between primary liver transplantation and death or final follow-up (March 1, 1989) and calculated by use of the Kaplan-Meier estimate. Retransplantation was disregarded because we were interested in patient, not graft, survival.

To calculate expected survival time for non-transplant patients the estimated aggregate non-transplantation survival function was computed (method available from author). Years alive with and without transplantation were calculated as the integral of the survival function, the numerical equivalent of the surface of the area below the survival curve. Two statistical tests, the one-sample log-rank and a simplified version of the actuarial prediction test,[16] were applied to the differences between observed and expected survival.

To examine the relation between the severity of pre-existing liver disease and the gain in survival with transplantation, the patient group was split into two approximately equal-sized subsets belonging to low (A) or high (B and C) Child-Pugh class.

Regression analysis was used to find out whether, at some time during the transplantation programme, there was an advance in the stage of disease at which patients were offered LTx.

The relation between the prognostic models was tested by pairwise and overall rank correlation of patient's prognostic indicators (Spearman's rank correlation and Kendall's coefficient of concordance).

Results

At the final follow-up date 19 of the 30 patients were alive (table 1). Death usually occurred within the first year after grafting. The wide differences between patients in clinical variables and prognostic indices reflect variations in hepatic dysfunction at the time of transplantation but, as expected, the subgroup with Child-Pugh class A had better values for the variables and indices than did patients in class B or C.

TABLE—PATIENT DATA

Group	All (n = 30)	Child-Pugh 5–7 (n = 12)	Child-Pugh 8–15 (n = 18)
*Characteristics**			
Albumin (g/l)	31·9 (6·2)	36·9 (4·5)	28·5 (4·8)
Bilirubin (mmol/l)†	216 (186)	102 (101)	292 (192)
Age (years)	49·6 (4·6)	47·8 (4·1)	50·8 (4·6)
*Prognostic indices**			
PI_c	5·63 (1·64)	4·33 (1·51)	6·49 (1·05)
PI_a	4·51 (2·25)	2·52 (0·92)	5·83 (1·86)
PI_m	7·45 (1·67)	5·84 (1·23)	8·52 (0·90)
Vital status			
Alive (at March 1, 1989)	19	7	12
Dead, surgical complication	3	1	2
Dead, hepatic artery pathology	4	3	1
Dead, cholestasis	4	1	3

*Findings given as mean (SD).
†Median: 223 (All), 45 (Child-Pugh 5–7), 279 (Child-Pugh 8–15).

Comparison of the observed and predicted survival probability curves indicate that LTx conferred a survival advantage from about 12 months after the operation, even though the three non-transplantation survival curves are not identical (fig 1). Cumulative survival probability with LTx at 1 year after transplantation was 0·65 (SE 0·09), which was similar to the Mayo prediction (0·63) but better than predictions from the Christensen or the AZG model (0·43 and 0·52 respectively). 5 years' survival with LTx evidently was superior to all predicted survival probabilities without LTx (0·58) versus 0·19 (Christensen), 0·29 (AZG), and 0·29 (Mayo). After 3 years every year of follow-up added about 0·3 years to expected survival gain per transplanted patient. The differences between observed and expected survival curves were significant in all three cases—by the one-sample log-rank test ($p = 0{\cdot}001$ [Christensen]; 0·014 [AZG]; 0·047 [Mayo]), and by the actuarial prediction test ($p < 0{\cdot}001$ [Christensen]; 0·002 [AZG]; 0·006 [Mayo]).

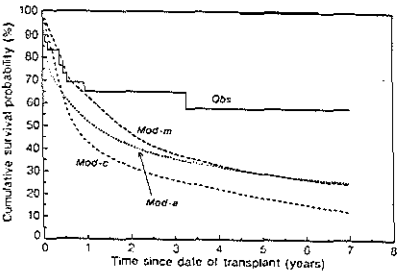


Fig 1—Observed and expected survival of PBC patients.

Obs = observed survival after LTx (Kaplan-Meier). Mod-c, Mod-a, Mod-m = predicted without LTx (from Christensen, AZG, and Mayo models).

Survival without transplantation for Child-Pugh class A patients (fig 2, upper) was close to that for survival after transplantation (p for difference from post-transplantation survival > 0·1 for all models), whereas patients in Child-Pugh classes B and C (fig 2, lower) clearly benefited from LTx ($p < 0{\cdot}001$ for each of the three models by log-rank χ^2 and Turnbull standard normal test statistics). The difference between these two subsets in survival gain is explained by the large differences in survival without transplantation; the subsets did not differ in survival after transplantation.

If the increasing experience gained by the transplant team was partly responsible for the apparent independence of transplantation survival from pretransplantation severity of disease, there should have been an increase in pretransplant disease severity with stage of transplant, without change in transplantation survival. Our data seem to support this supposition. Survival curves for the 15 patients who received transplants in the first half of the programme were the same as for the 15 in the second half, even though prognostic indices were higher for those entered later in the programme, irrespective of the PI chosen. Also the SD fell, which suggested a decreasing variation in the stage of the disease at which the patient underwent transplantation.

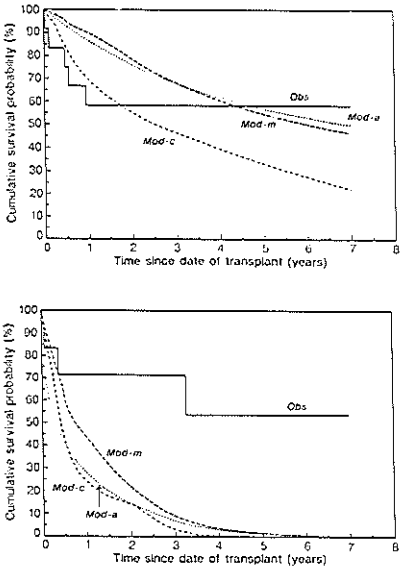


Fig 2—Observed and expected survival of PBC patients according to Child-Pugh score.

Upper: Child-Pugh class A.
Lower: Child-Pugh classes B and C.

Regression analysis with the three prognostic indices for each patient as the dependent variable and LTx-serial number as the independent variable showed that, whichever of the three models used, prognostic indices were greater the later the patient entered the LTx programme; this was presumably due to overrepresentation of patients with relatively good liver function at the start of the programme (fig 3).

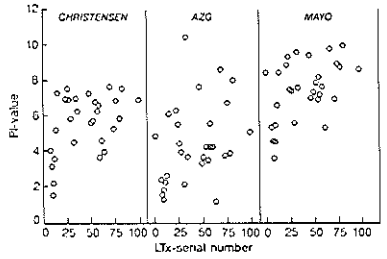


Fig 3—Prognostic indicators related to LTx serial number.

Most results show a firm relation between prognostic models. We investigate the relation between the prognostic indices as a test for validity of the prognostic models. The Child-Pugh score (ranging from 5 to 15) was included as a separate prognostic indicator (not index). Only rank correlation statistics may be used because absolute survival probabilities are not directly represented by the PI or the Child-Pugh score. Pairwise comparison of the four prognostic indicators resulted in highly significant rank correlations, all exceeding $r_s = 0{\cdot}80$. The overall agreement of prognostic indicators was high, Kendall's coefficient of concordance being 0·93 ($p < 0{\cdot}001$).

Discussion

Our findings show that LTx significantly improves the long-term survival of patients with PBC of Child-Pugh classes B and C. Although LTx may sometimes be justified for improving the quality of life for patients with class A primary biliary cirrhosis, it will probably not lengthen survival. Predictions beyond 7 years are limited by both the time horizon of the prognostic models (survival without transplantation) and by the limited follow-up time for transplant patients (transplantation survival). Overall the conclusions are in agreement with those of Neuberger[20] and Markus,[21] although their studies differ from ours in the number of prognostic models used, the statistical method of aggregation of individual prognostic information, and the patient group.

Neuberger's and Markus' studies each apply only one prognostic model, the Christensen-model and the Mayo-model, respectively. A disadvantage of the use of these models for prediction of survival without transplantation is their derivation from a patient population in an early stage of the disease. Thus, application of these models requires extrapolation to more severe stages, which could lead to biased estimates. The AZG-model was added to counter this uncertainty, but the limited number of patients on which the final model was based ($n = 76$) means that its statistical reliability is poorer than that of the Christensen and Mayo models. However, the validity may be better because, compared with the patients on which the Christensen and Mayo models are based, those for the AZG model are more recent (year of entrance from 1978 to 1985) and have a distribution of severity of disease (Child-Pugh A 54%, Child-Pugh B 27%, Child-Pugh C 19%) that encompasses that of transplanted patients.

Another difference concerns the aggregation technique applied in the respective studies. To arrive at a curve for group survival without transplantation for comparison with the Kaplan-Meier transplantation survival curve Neuberger computed the arithmetical mean of the individual prognostic indices of his 29 patients. The corresponding non-transplantation survival function was assumed to reflect the group's survival without transplantation. This assumption is not justified from a mathematical point of view, since particular prognostic indices for individual patients have an exponential relation to the survival function. Unpredictable bias will result (fig 4).

The aggregation technique of Markus implies averaging all individual predicted survival curves. This procedure may be justified when there is little variation between individual prognostic indices over time. When non-random time-dependent variation exists, the comparison with Kaplan-Meier estimates of transplantation survival may be seriously biased, since patients with recent transplants have only short-term influence on transplantation survival but

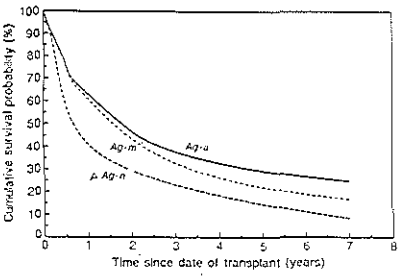


Fig 4—Aggregated predicted survival curves. Example with Mayo-model and all AZG-patients (n = 30).

Ag-n, Ag-a, Ag-m = aggregation by Neuberger, AZG, and Mayo, respectively.

full-term influence on estimates of survival with transplantation. If, for example, severity of pretransplantation status increases over time, the gain of survival with LTx will be overestimated (fig 4).

Finally we compared the balance of transplantation survival and survival without transplantation in the AZG with Neuberger's result (PBC patients who received transplants in Cambridge, England, until April 1984) and with those of Markus et al (161 PBC-patients who received transplants in Pittsburgh and the Mayo-clinic between March, 1980, and June, 1987).

Neuberger and colleagues' patients were generally at a later stage of the disease (average PI_e 6·76 [SD 0·75]) than were the AZG-patients (average PI_e 5·63 [1·64]); other indices were not available. Survival after transplantation (especially in the peri-operative period) and also the estimated gain in survival with transplantation was smaller than that for the AZG-patients in Child-Pugh classes B and C. This finding suggests a limit to which the operation can be postponed.

Our results for short-term transplantation survival seem inferior to those of Markus and colleagues, who report virtually no peri-operative deaths. This discrepancy may be due to differences in definitions of patients included in the analysis, or it may reflect the greater experience of the American centres.

The consistency between the three independent models in predicting survival without transplantation have both methodological and clinical implications. They show that a specific application of Cox' proportional hazards model is able to provide predicted survival curves which may be validly compared with observed Kaplan-Meier survival curves. Many applications of the above described procedure are conceivable.

This study was financed by the Dutch National Health Insurance Board (Ziekenfondsraad). We thank Mrs Marie-Louise Bot for help with data collection and preparation of the paper.

REFERENCES

1. Maddrey WC, Thiel DH van. Liver transplantation: an overview. *Hepatology* 1988; 8: 948–49.
2. Starzl TE, Demeter AJ, Thiel D van. Liver transplantation (first of two parts). *N Engl J Med* 1989; 321: 1014–22.
3. Starzl TE, Demeter AJ, Thiel D van. Liver transplantation (second of two parts). *N Engl J Med* 1989; 321: 1092–99.
4. O'Grady JC, Williams R. Present position of liver transplantation and its impact on hepatological practice. *Gut* 1988; 29: 566–70.
5. Cuthbert JA. Southwestern Internal Medicine Conference: hepatic transplantation. *Am J Med Sci* 1986; 291: 286–96.
6. Krom RAF, Gips CH, Newton D, et al. A successful start of a liver transplantation program. *Transpl Proc* 1983; 15: 1276–78.
7. Scharschmidt BF. Human liver transplantation: analysis of data on 540 patients from four centers. *Hepatology* 1984; 4: 95S–101S.
8. Bonsel GJ, Habbema JDF, Bot ML, et al. A technology assessment of liver transplantation. An evaluation of the Groningen liver transplantation program 1977–1987 (in Dutch). *Ned T Geneeskd* 1989; 133: 1406–14.
9. Bonsel GJ, Klompmaker IJ, Bot ML, et al. Cost-effectiveness analysis of the Dutch liver transplantation programme. *Transpl Proc* (in press).
10. Bismuth H, Erlezon BG, Rolles K, et al. Hepatic transplantation in Europe: first report of the European Liver Transplant Registry. *Lancet* 1987; ii: 674–76.
11. Pugh RNH, Murray-Lyon IM, Dawson JL, et al. Transection of oesophagus for bleeding oesophageal varices. *Br J Surg* 1973; 60: 646–49.
12. Conn HO. A peek at the Child-Turcotte classification (editorial). *Hepatology* 1981; 1: 673–76.
13. Putten ABMM van der, Bijleveld CMA, Slooff MJH, et al. Selection criteria and decisions in 375 patients with liver disease, considered for liver transplantation during 1977–1985. *Liver* 1987; 7: 84–90.
14. Slooff MJH, Bams JL, Sluiter WJ, et al. A modified cannulation technique for veno-venous bypass during orthotopic liver transplantation. *Transpl Proc* 1989; 21: 2328–29.
15. Klompmaker IJ, Haagsma EB, Gouw ASH, et al. Azathioprine and prednisolone immunosuppression versus maintenance triple therapy including ciclosporine A for orthotopic liver transplantation (OLT). *Transplantation* (in press).
16. Christensen E, Neuberger J, Crowe J, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis. Final results of an international trial. *Gastroenterology* 1985; 88: 156–65.
17. Bonsel GJ, Veer F van 't, Bot ML. Costs and effects of liver transplantion, vol 4: Prognosis of liver transplantation, the Groningen program 1977–1987 (in Dutch). Rotterdam: Erasmus University Rotterdam, 1988.
18. Dickson ER, Grambsch PM, Fleming TR, et al. Prognosis in primary biliary cirrhosis: model for decision making. *Hepatology* 1989; 10: 1–7.
19. Buxton M, Acheson R, Caine N, et al. Costs and benefits of the heart transplant programs at Harefield and Papworth hospitals, Final report, page 171–2. Brunel University, University of Cambridge, 1985.
20. Neuberger J, Altman DG, Christensen E, et al. Use of a prognostic index in evaluation of liver transplantation for primary biliary cirrhosis. *Transplantation* 1986; 41: 713–16.
21. Markus BH, Dickson ER, Grambsch PM, et al. Transplantation improves survival in patients with primary biliary cirrhosis: comparison of estimated survival based on Mayo model and actual survival in the Pittsburgh transplant population. *N Engl J Med* 1989; 320: 1709–13.

Quality of life measurement and medical technology assessment

Gouke J. Bonsel

Appeared in Dutch in: Nederlands Tijdschrift voor de Psychologie 1991; 46: 180-186)
Published with permission of the editor

Abstract

Medical technology assessment (MTA) is a type of policy research which supports decisions at the macro-level on the allowance of medical technologies at the health care market. MTA puts a high value on the quantitative analysis of the multiple effects of medical technologies on patients. Analysis of survival and quality of life (QOL) are the core of this outcome analysis.
In this context QOL methodology consists of methods for description and methods for valuation. These methods share the application of standard psychometric techniques for operationalization and measurement. An example of description and valuation is drawn from the Dutch MTA of heart transplantation.
Criticism on QOL-research may partly be explained by its developmental stage and by the cultural differences between the three disciplines involved.

1 Introduction

Psychology is currently an acknowledged discipline within clinical medicine, and psychologists are to be found in many positions within the medical system. Its area of application is still growing and is not restricted to mental health care alone. Clinical psychologists, but other professionals too, routinely apply psychological methods of diagnosis and treatment in every day medical care. Additionally, clinical research currently includes the use of psychological methods of data collection (structured interviewing, questionnaires) and analysis. This paper on quality-of-life research (QOL-research) presents a recent example of this general development.
Generally speaking, QOL-research is directed towards the measurement of health status. Depending on the particular nature of the research, health status may be operationalized as the perceived well-being of the patient, as rather objective characteristics of health itself, or as specific or global determinants of (perceived) health. The boundaries of QOL-research are rather vague, and theoretical and empirical problems are abundant. The latter for example include the definition of 'health' and the measurement of QOL in severely diseased patients.
Last decades has showed an impressive growth of QOL-research (Knippenberg, 1988), particularly in two areas: the medical treatment and support of cancer patients (decisions on the micro level) and the societal decision-making on the

introduction of new medical technologies (decisions on the macro level). The latter policy research, which is usually initiated by the national government or health care authorities, has been denounced as medical technology assessment (MTA) (Bonsel 1990).
This contribution to of the Dutch Journal of Psychology will be dedicated to QOL-research in the context of MTA and therefore on decisions on the macro level (Habbema, 1990). We start with some theoretical considerations on MTA and QOL, followed by the presentation of a case-study which illustrates the theoretical and empirical approach. Finally a (partial) reply will be offered to critical questions which has been rised by both investigators in the field and people involved outside.

2 Medical technology assessment

MTA systematically investigates the consequences of the application of a medical technology from a societal point of view.
In this context a comprehensive definition of *medical technology* is adopted, including pharmaceuticals, diagnostic imaging devices, surgical procedures and organizational provisions. One, or rather a complex of technological innovations may be concerned, which is applied within a particular health care programme dedicated to patients with a specific disease or health-related problem. Thus an MTA of heart transplantation is not restricted to the cardiac surgery alone, but also includes the extensive health care programme for these type of patients (usually suffering from severe heart failure). The surgery is only a modest part of this programme. The care which is offered after the transplantation operation, and offered to those rejected or awaiting a donor organ, they all are subject to an MTA of heart transplantation. Thus an MTA and the related policy decisions have to be placed in a wide context: not a single medical service but a comprehensive health care programme is at issue.
In MTA *consequences* too are rather broadly defined. They encompass desirable and undesirable effects on patient outcome, effects on patients which are not accepted for the technology under investigation, and they also include the employment of resources, the impact on the health care system at large, and the ethical and legal consequences. Those consequences with refer to the perceived well-being

of the patient and his performance are designated as quality-of-life effects[1].
Systematically points to a. the comprehensiveness of the consequences, b. the formal synthesis of empirical results (multi-criteria analysis, cost-effectiveness analysis, cost-benefit analysis)[2], and c. the generalizability of conclusions from the empirical research.
The *societal viewpoint* is represented by many specific features of the design and methods of investigation (Habbema, 1989). One particular aspect of the societal viewpoint is the emphasis on the analysis of the health effects in general, rather than on the analysis of biomedical parameters (like the pumping capacity of the heart in heart transplantation patients). Another specific feature of outcomes measurement in MTA in this respect is the valuation of health outcomes additionally to their description. Descriptive methodology is rooted in psychology (scaling techniques, classical test theory). Methods for valuation are derived from economic theory (on the utility of commodities) and psychometric decision theory. An elaboration is presented in the next paragraphs.

3 Concepts in the description and the valuation of health status

The point of departure in any type of QOL-research is the choice of a particular *concept of health*. If we abandon a negative definition of health[3], then various strategies exist to define the concept of health:

1. a philosophical strategy: according to a particular religion or conviction, the (most) important dimensions of health may be defined. One example is the WHO-concept of health, which distinguishes between a physical, a mental and a social dimension (Breslow, 1972);
2. a democratic strategy: according to the results of population surveys, those dimensions are selected which are regarded to be essential for health. Several

---

[1] If the emphasis is on performance and other 'objective' characteristics, we prefer the designation 'health status' instead of 'quality of life'. (In this paper quality-of-life and health status as a rule are regarded to be synonyms).

[2] Apparently some form of standardisation is needed.

[3] I.e. health as the (degree of) absence of disease.

health status questionnaires have been devised according to this strategy (e.g. The Notthingham Health Profile; Hunt, 1986).

3. a discriminative strategy: according to prior knowledge on the specific effects of a technology, those dimensions are chosen which are expected to record the largest changes. Usually this strategy restricts the attention to one dimension or aspect which is strongly associated to a specific biomedical parameter (e.g. the aspect 'body image' if a psychological mutilating surgical procedure is compared to a more conservative one).

Regardless the strategy, all concepts share the premise that health is a theoretical construct which should be operationalized[4], rather than an empirical notion, which can be measured straight forwardly. If the choice of the health concept is explicitly accounted for, the first strategy appears to be the more common one, though the result often will resemble that of the second strategy. Usually the construct health is assumed to be covered by different 'aspects' or, in a non-statistical meaning by, 'dimensions'. Frequently, the dimensions of the WHO-definition constitute the nucleus, supplemented with, for example, the dimension work.

The *operationalization of the dimensions* makes up for a second stage of decision-making on the design of QOL-research. The operationalization may be accompanied by preparatory investigations. After this stage, we are able to classify and consequently to describe health status or changes in health status.

Being aware that the description of health status takes with it some arbitrary - but not random - choices, we admit that the *valuation* of health status requires even more decisions to be made. This part of QOL-research aims at assigning value(s) to health status in an economic interpretation of the word 'value'. The ultimate goal of this enterprise is to relate societal investments directed at health improvement (as expressed in monetary units), to the value of changes in health which are actually observed (as expressed in units yet to be defined). See for details Torrance (1986).

---

[4] Here we will not elaborate on theories of operationalization and particularly not on the relevance of a nomological network.

One might expect that the subjective judgment of the patient with regard to his own health or health improvement, may suffice for such a valuation[5]. However, despite its occasional application in MTA, this method may be challenged fundamentally. The perceived well-being depends on many circumstantial factors. The majority of these factors is unrelated to the actual level of health, even in diseased persons. Particularly we should mention the influence of the personal set of reference values and the potential effect of 'coping' on patient judgments of their own well-being. If we intend to apply the results in societal decision-making on medical technologies, these health-unrelated factors should be regarded as confounders (Drummond, 1987; Campbell, 1976). Biased responses may also occur if a patient uses his response as a messenger for promotion of the treatment or as a mean to express his discontent with the doctor or his treatment. Thus the use of subjective evaluations of the patients in MTA is limited[6].

An example, which is taken from the MTA of liver transplantation, may support the argument (Bonsel, 1989). In this study patients, who were alive three months after transplantation, reported a level of experienced well-being which exceeded that of the normal population. In due time this level attenuated, despite the fact that the medical condition was still improving. Feelings of gratitude and relief for the provisional success of this risky procedure apparently rivalled the role of actual medical improvement. Experienced well-being thus appears to be an invalid measure of outcome in economic analysis in the context of societal decision-making.

After this argument on the relative unsuitability of patient's valuation of his own health, we proceed with an account of methods thought to be more appropriate. They all are *explicit* valuation techniques, which are derived from economic and decision analytic theory. (Fanshel, 1970; Torrance, 1976A & 1976B, Pliskin, 1980). These techniques explicitly elicit value judgments on health status, either in a direct or an indirect way. The direct way offers the respondent one or more response modalities to value health states (Kaplan, 1979). The indirect way asks the respondent to indicate his behaviour in a projected health situation.

---

[5] We prefer to define a change in health (status) as the difference of two health (status) measurements rather than one measurement post hoc of the difference as such.

[6] Their use as a support to individual decisions.

The respondent should choose between either certainty of no change in a suboptimal health state or a probability of improvement with a complementary probability of worsening or death. Patients themselves are not primarily engaged in this valuation process.
The description and valuation of health status in the context of an MTA will be illustrated with an example in the next two paragraphs. The example is drawn from the Dutch MTA of heart transplantation which was carried out between 1985 and 1988. This MTA was instituted to provide answers on many questions about heart transplantation, for example the effects on patient's survival and quality of life, the costs and the need for heart transplantation in view of limited donor organ availability (Charro, 1988). At this place we confine ourselves to the investigation of the quality of life effects of health transplantation (Bonsel, 1988).

4 The Dutch MTA of heart transplantation: the description of quality of life effects

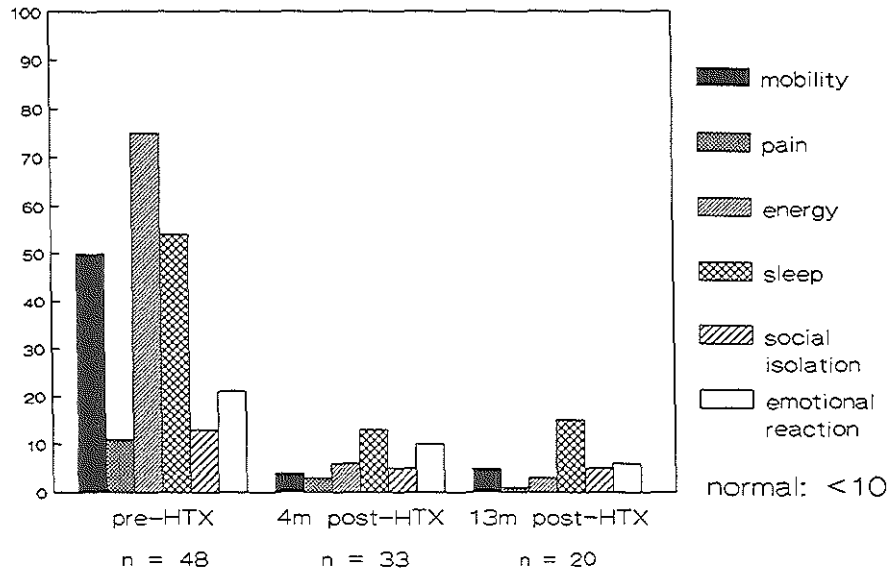
As a randomized controlled clinical trial appeared to be ethically unacceptable and unfeasible for practical reasons, the study design was observational. A control group of patients was lacking. For analytical purposes we assumed that in absence of transplantation quality of life of patients would remain stable at best.
As a concept of health we adopted the WHO-definition. Standard-questionnaires were selected on psychological functioning (State Trait Anxiety Inventory [Ploeg, 1980], Self-Rating Depression Scale [Zung, 1965], Psychological Questionnaire for Patients with Heart Disease [Erdman, 1982]) physical functioning (questions derived from the Dutch National Health Survey, disease-specific questionnaires), social functioning, work and sexual functioning.
As a separate issue experienced well-being was measured on a global and a specific level, applying the Index of Well-Being (Campbell, 1976) and Life Satisfaction questions from the Dutch National Study on Living Conditions respectively.
It appeared that the number of QOL-measurements and the precise timing of the interview implied a choice on the conceptual level. Three considerations will be presented. If an experimental design is adopted in clinical research, the measurement instruments and the measurement schedule are deliberately chosen in order to maximize the probability to demonstrate differences between alternative actions.

In treatment evaluation the resulting schedule usually consists of two measurements, one just prior to a treatment cycle and one shortly afterwards. This type of schedule may induce statistically significant differences without clinical or societal relevance in the long run. However, as MTA should support far reaching decisions at the macro level, due attention should be paid to ultimate effectiveness rather than to temporary or short term effects.
The second conceptual consideration on the time schedule follows from the phenomenon that the changes in each dimension only infrequently coincide. For example in heart transplantation we observed a resumption of elementary daily activities (washing, dressing, etc.) within a week after transplantation. Psychological recovery at least took some weeks, and the social recovery process and the restoration of full employment even lasting months or longer.
The last consideration takes the length of treatment into account, which in many cases is life-long. In the example of organ transplantation this is clearly visible by the permanent immuno-suppressive treatment following transplantation.
For these reasons in the heart transplantation study a longitudinal design was adopted resulting in quarterly interviews after transplantation and one interview just prior to transplantation.
Table 1 and figure 1 show some results from the descriptive QOL-analysis of this study. The table shows average scores from questionnaires on performance status, anxiety, depression, experienced well-being and work. Results represent the health state at three important points in time: just prior to transplantation, 4 months after transplantation and 13 months after transplantation. Four months after transplantation a dramatic improvement may be observed. Further improvement, though present, is rather modest. The depression-scores and other data (not shown) demonstrate the presence of psychosocial problems. Due to intensive professional support these problems usually appeared to be temporary (Bonsel, 1990). Multivariate analysis did show age, diagnosis, gender and education not to influence results significantly. If this influence had been demonstrated - as was the case in an analogue American study - the societally relevant question might have been raised whether the explaining variable(s) should be taken into account during the selection of patients. (Selection is inevitable, due to the discrepancy between the availability of donor organs and the prevalence of patients with end-stage heart failure).
This type of consideration is a specific feature of MTA: many choices, which otherwise should be made from an academic point of view, should de made prudently.

Table 1 Quality of life before and following heart transplantation

Questionnaire	Normal	Before Tx	4 Mnths Post Tx	13 Mnths Post Tx
Karnofsky Performance Index	> 90	53	89	94
State-trait Anxiety Inventory (STAI-Spielberger)	< 37	46	35	29
Self-rating Depression Scale (SDS-Zung)	< 38	55	43	41
Index of Well-being	> 12	7	13	14
Work/Regular Activities if $\geq$ 2 hr/d	? > 90%	23%	60%	63%

Figure 1 Nottingham Health Profile before and following heart transplantation



5 The Dutch heart transplantation MTA: the valuation of quality of life effects

The following account of valuation of quality of life effects distinguishes between four stages. Each stage will be supported by some empirical evidence from the transplantation study. We assume the effect of - in our case heart transplantation - on survival is known, which implies knowledge on the survival probabilities with and without transplantation.

Stage 1 The health state of patients is measured empirically by means of standard questionnaires (our choices have been described in the previous paragraph). A longitudinal design is adopted, accounting for nine measurements, one before and eight at quarterly intervals following transplantation.

Stage 2 For each interval 'health scenario's' are derived from these empirical data by means of a formal algorithm. A health scenario consists of a standardized description of a health state. In this case four dimensions were regarded to be essential for such a description: mobility/self-care, psychological state (anxiety, depression), symptoms/complaints (loss of energy, shortness of breath) and intensity of medical care.

Figure 2 An example of a four-dimensional health scenario

Somebody

- is at home
- is rather quickly short of breath, at times in pain
- is able to care for himself, though limited in walking around
- has no psychological complaints and having trust in the future

---

Figure 2 is an example of a health scenario based on our four dimensional system. If the questionnaire data did approve of heterogeneity of health states within a particular period, more than one health scenario was abstracted from the data. In our study this procedure yielded well over 20 health scenario's.

Stage 3 A value is attached to each health scenario by means of a specific valuation method. Usually the standard-gamble method is regarded to be the superior method from a theoretical point of view[7]. However, the operationalization of the standard-gamble method is hampered by several problems. A simpler method is preferable, e.g. visual analogue scaling (VAS), provided a fixed, mathematically defined relation may be assumed between the VAS values and the standard-gamble values. In our study the about 20 health scenario's were presented in a random order to a panel of laymen. The VAS-method was applied. Despite the apparent complexity of the valuation task, the panel ranked and valued the health scenario's quite unanimously. Next, values were indexed using a 0-1 scale and the average value of each health scenario was calculated. Finally we arrived at a value for each interval or period by weighted averaging of health scenario values. The weights were derived from the frequencies of the particular health states in a given period as observed in Stage 1. In this example the pre-transplantation period value was about 0.35, the first average value in the year after transplantation was about 0.75 and the subsequent year this value was about 0.9.

Stage 4 Survival consequences were adjusted with the above-mentioned values, resulting in so called Quality Adjusted Life Years (QALY's) as measure of (economic) outcome. Once again the heart transplantation study example: if we assume survival to be on average three months without transplantation and ten years with transplantation, and if we assume the health state to remain stable after two years, the next benefit of heart transplantation may be calculated as follows: $(0.75 \times 1 + 0.9 \times 9) - (0.25 \times 0.35) = 8.76$ QALY's. This outcome may be formally related to net costs in a so called cost-utility analysis (De Charro, 1988).

---

[7] Essentially standard gamble compares two health states, A and B respectively. The respondent is offered the following alternatives. Alternative 1 includes certainly (probability = 1.0) that the respondents remains in (suboptimal) health state A for the remaining part of his life. Alternative 2 offers a treatment modality for health state A with two possible outcomes: either the patient recovers to full and permanent health (health state B) which has a probability of Q, or the patient dies instantaneously with probability 1-8. The valuation process consists of an iterative procedure which varies probability Q, until the respondent expresses indifference to the two alternatives. The utility of health scenario A follows from the following equation: $U_A = Q^* U_B + (1-Q)^* U_{death}$. U_B is usually set to 1, U death to 0. Note that reference health scenario's U_B and U death may be chosen otherwise provided that $U_B \geq U_B \geq U_{death}$ holds.

6 The position of QOL-research in MTA: internal and external criticism

Obviously the analysis of effectiveness to which QOL-research belongs, is the nucleus of most MTA's. The balance of positive and negative effects frequently is the fundamental issue in health policy decisions at the macro level (like it is in individualized decisions on the micro level). This key position does not protect us against many practical problems (in the example the absence of a control group, the shortness of follow-up, the selective losses to follow-up), nor does it preserve us from theoretical problems. The latter type of problems may be distinguished into tensions between scientific cultures ('criticism from outside') and problems due to unsolved theoretical questions ('criticism from inside').

6.1 Criticism from outside

The analysis of effectiveness brings the independent MTA-researcher and the clinical investigator together. Usually part of the effectiveness analyses within the context of a MTA does not appeal to the clinical investigator, despite a sincere interest in clinical outcome of the latter. However, the objectives of effectiveness analysis, though related, are not identical and the research environment frequently is quite different. Some examples of these differences follow.

1. From the viewpoint of the MTA-investigator the health care programme is the *unit of analysis*, from that of the clinician it is the patient treated. Consequentially, differences may be observed with regard to the attention paid to a. non-treated patients who are involved in the programme, and b. to the patient's environment. The latter difference points to a question yet to be solved, i.e. whether the current primacy of patient's interests in individual decisions necessarily implies the neglect of effects on patient's environment if decisions on a health programme at large are concerned.
2. The MTA-investigator prefers disease-non-specific, generally applicable scales as *measurement instruments*. The emphasis is on the level of scores and on the amount change observed. These are related to levels and ranges which are known from the general population. The MTA-investigator collects patient's information on the process of medical care too. As a contrast the clinician prefers medical parameters as a measure of outcome, occasionally supplemented

with disease-specific questionnaires. Apparently the main interest is in showing a significance between the target intervention and some alternative at a given point in time. Process information usually is not regarded to be scientifically interesting.

3. Differences are not confined to the specific study objectives and the set of outcome parameters chosen. *Additional methodological differences* do exist. The MTA-investigator applies the full range of methods and techniques which are regarded standard tools within the social sciences (for example scaling techniques). The clinician prefers medically based techniques, which do not depend (much) on the 'unreliable' response of the patient. Unobservable variables and a high measurement level which is achieved by the application of scaling techniques, are distrusted at best (Meer, 1988). The request from the heart transplantation centres to skip all the superfluous items from the SDS, STAI and other questionnaires and to select 'a few meaningful questions' is a good example of these intercultural differences.
4. Different views on the *time schedule* of measurement provide us with a last example of these differences.

6.2 Criticism from inside

This paper on the position of QOL-research within MTA finishes with an indication of acknowledged scientific problems in this field. According to some authors these problems seriously affect the position of QOL-research in the context of health policy decisions on the macro level, resembling discussions which dominated the social sciences for tens of years.

The first problem in the context of health status *description* is the lack of consensus on a standard repertoire of generally applicable, disease-non-specific questionnaires. An important ruling principle is the selection of originally English-language questionnaires. This principle, which dates back to the strong incentive to publish in English-language scientific journals, severely curtails the opportunity to apply additional scientific criteria for the selection of questionnaires. A second problem is the limited availability of reference data from the normal population, even if the source version has not been constructed elsewhere.

A third barrier to comparability of results is caused by the presence of rather inflexible local research policies and by the silent influence of substantial economic interests. Even in a small country like the Netherlands, standardization appears to be a long way going.
More complex problems are encountered in the *valuation* of health status. This hybrid of medical, psychological and economic theory is bristling of pitfalls. The assumption that a QALY conceptually may be regarded as a measure of utility, theoretically implies empirical observations which yet have to be established. We especially think of the independence of the adjustment index and the time dimension (Torrance 1976A).
The multitude of valuation methods makes up for a second theoretical challenge. We already mentioned two of these methods, standard gamble and visual analogue scaling. All methods share that little is known on the internal validity, and even less on the external validity. Additionally, feasibility and validity are counteracting criteria which necessarily require a compromise of design features of the Stages 2 and 3. This compromise is by no means standard.
Thirdly, the aggregation of individual utilities to arrive at a substitute for society's utility has been disputed. Despite their complexity, a collaborative group of Dutch and other European investigators has demonstrated that the latter two problems may be solved. Continued integration of medical, psychological and economical theories is a prerequisite for this progress, which in our view could benefit even more from existing psychometric evidence (Kind, 1982).

7 Epilogue

This paper has provided an outline on a research area in medicine which owes much to psychological theory. This area, indicated by 'QOL-research', concerns the description and the valuation of (changes in) health state. The apparently growing interest in QOL-research is caused by the acknowledged contribution of psychological experience and theory to the individualized treatment of patients - e.g. suffering from cancer - who face trade-offs between quality-of-life and survival. This contribution currently is extended to the field of decisions on medical technologies on the macro-level, particularly by means of its undeniable influence on QOL-research in MTA.

The challenge of QOL-research to any scientist regardless his academic background and his main interest - empirical or theoretical - in our view is evident. Beyond doubt further integration of social sciences in this type of research is relevant for future progress. We expect that this integration will enhance the arousing awareness of clinical scientists of the usefulness of this body of knowledge to medical research. Current developments in many medical specialties (besides oncology we think of rheumatology, cardiology, surgery and primary care) support this expectation.

References

Bonsel GJ, Bot ML, Veer F van 't. De kosten en effecten van harttransplantatie, deelrapport 2C: De kwaliteit van leven voor en na harttransplantatie: Resultaten. Rotterdam: Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit, 1988.

Bonsel GJ, Erdman RAM, Mast RC van der, Balk AHMM, Maas PJ van der. Psychosociale aspecten van harttransplantatie - 4 jaar ervaring. Nederlands Tijdschrift voor Geneeskunde 1990; 134, 227-231.

Bonsel GJ, Habbema JDF, Bot ML, Veer F van 't, Charro FT de, Maas PJ van der. Een "technology assessment" van levertransplantatie; een onderzoek naar het Groningse levertransplantatieprogramma 1977-1987. Ned Tijdsch Geneeskd 1989; 133, 1406-14.

Bonsel GJ, Rutten FFH. Goede raad is duur? Medisch Contact 1990; 45, 512-516.

Breslow L. A quantitative approach to the World Health Organization definition of health: physical, mental and social well-being. Int J Epidemiol 1972; 1, 347-55.

Campbell A. Subjective measures of well-being. New York, Plenum Press, 1976.

Campbell A, Converse PE, Rodgers WL. The quality of American life. New York, Russel Sage Foundation, 1976.

Charro FT de, Bonsel GJ, Hout BA van. De kosten en effecten van harttransplantatie, deelrapport 9: Eindrapport. Rotterdam: Vakgroep Economie, Juridische Faculteit, Erasmus Universiteit, 1988.

Drummond MF. Resource allocation decisions in health care: a role for quality of life assessments? J Chronic Dis 1987; 40, 605-16.

Erdman RAM. Een medisch psychologische vragenlijst ter bepaling van het welbevinden bij hartpatiënten. Hart Bulletin 1982; 13, 143-7.

Fanshel S, Bush JW. A health status index and its application to health-services outcomes. Operations Research 1970; 18, 1021-66.

Habbema JDF, Casparie AF, Mulder JH, Rutten FFH. Medische Technology Assessment en gezondheidsbeleid. Alphen aan de Rijn, Samson Stafleu, 1989.

Hunt SM, Mc Ewen S, McKenna SP. Measuring health status. London, Croom Helm, 1986.

Kaplan RM, Bush JW, Berry CC. Health status index. Category rating versus magnitude estimation for measuring levels of well-being. Medical Care 1979; 17, 501-25.

Kind P, Rosser R, Williams A. Valuation of quality of life: some psychometric evidence. In: Jones-Lee MW. The value of life and safety. Amsterdam, New York, Oxford: North Holland Publishing Company, 1982.

Knippenberg FCE van, Haes JCJM de. Measuring the quality of life of cancer patients, psychometric properties of instruments. J Clin Epidemiol 1988; 41, 1043-53.

Meer J van der. Moeilijke keuzen aan het ziekbed; reken niet op de besliskunde. Ned Tijdsch Geneeskd 1988; 132, 336-7.

Pliskin JS, Shepard DS, Weinstein MS. Utility functions for life years and health status utility functions. Operations Research 1980; 28, 206-24.

Ploeg, HM van der. Validatie van de zelf-beoordelingsvragenlijst. Ned Tijdsch Psychol 1980; 35, 243-49.

Till JE, McNeill BJ, Bush RS. Measurements of multiple components of quality of life. Cancer Treatment Symposia 1984; 1, 177-81.

Torrance GW. Measurement of health utilities for economic appraisal. A review. J Health Economics 1986; 5, 1-30.

Torrance GW. Health status index models: a unified mathematical view. Management Science 1976A; 22, 990-1001.

Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. Socio-economic Planning Science 1976B; 10, 129-36.

Zung WWK. A self-rating depression scale. Arch Gen Psychiatry 1965; 13, 63-70.

Assessment of quality of life before and following liver transplantation: first results

Gouke J. Bonsel
Marie-Louise Essink-Bot
Ids J. Klompmaker
Maarten J.H. Slooff

Accepted for publication in Transplantation

Abstract

Analysis of the quality-of-life effects is part of the current technology assessment of the Dutch orthotopic liver transplant-programme.
Data are collected by means of computer-assisted interviewing, including one interview before transplantation and annual follow-up interviews. Data on psychiatric morbidity are obtained from medical records.
This article shows preliminary results of a cross-sectional analysis of data collected from 1987 to 1989. Eighty-eight measurements were obtained from 46 adult patients (response rate 82 %).
Pre-transplant scores suggest major restrictions on all domains of life, especially a low amount of energy. After transplantation all indicators show improvement, although the level of the general population is not always attained. Improvement of subjective quality of life is more marked, maybe due to euphoria at surviving the hazardous procedure. Psychiatric events occurred only infrequently.
We conclude that orthotopic liver transplantation contributes positively to the quality of life of surviving patients. In addittion, computer-assisted interviewing proved to be a feasible survey technique, even when very ill patients were involved.

1 Introduction

Though never assessed in a randomized clinical trial, orthotopic liver transplantation is accepted as an effective therapy for patients with end-stage liver disease (1). The considerable claim on financial resources and the uncertainty about the balance of harm and benefit have given rise to many investigations of its overall effectiveness. Results on the gain in length of survival of patients treated with liver transplantation have recently become available (2-4).
However, studies assessing the gain in quality of life as a result of liver transplantation are few (5-6). Most studies on the effect of liver transplantation on the quality of life have studied a few non-comprehensive indicators of quality of life, e.g. days spent in hospital, work, activity status, and growth in length (children) (7-13). In addition several reports are available on psychiatric morbidity, focusing primarily on neuropsychiatric symptoms of end-stage liver disease (hepatic encephalopathy) (14-16). A more comprehensive assessment of quality of life in liver

transplantation patients was reported for liver transplant recipients only after the operation, probably due to considerable difficulties in obtaining measurements from pretransplant patients (17-20). The study of Lowe (20) is of particular reference since it applied a quality of life instrument (the Nottingham Health Profile) which was also used in our study. Only the study published by Tarter in 1988 has a longitudinal design (21); preliminary results indicated a sharp improvement in quality of life after liver transplantation, although the pre-morbid level of quality of life usually was not attained. The present article describes the design and the first results of a study on the changes in health-related quality of life in adult Dutch liver transplant patients. The study is part of a medical technology assessment of the liver transplantation programme in the Academisch Ziekenhuis Groningen in the Netherlands, which analyzed effects of liver transplantation on survival, quality of life and various resources (3,22).

2 Patients and methods

2.1 Patients

The protocol of patient selection, timing of transplantation and support during follow-up has been described before (23-25). The nationally accepted protocol states overt psychiatric morbidity, including active alcoholism, to be a contra-indication (3). Intensive support by a specialized social worker is a regular part of the clinical programme.

A longitudinal study design for the quality of life study with pre- and post transplant measurements was aimed at. Until the start of data collection for the quality of life study in June 1987, 63 patients received a transplant, 38 of whom were still alive 1 month to 8 years post-transplantation. These 38 patients constituted the group that was only eligible for post-transplant measurements (the cross-sectional group), for obtaining information on the long-term quality of life. Eight patients under 18 were omitted; additionally 4 adults were omitted for practical reasons as they lived abroad. The remaining 26 Dutch adults living with a transplant in June 1987 were included in our study and completed a questionnaire once a year.

The longitudinal group consisted of all adult Dutch-speaking liver transplantation candidates who have entered the programme since June 1987 as defined by the formal request for a donor liver from Eurotransplant. Collection of data for the present report was terminated on July 1st, 1989. By then the longitudinal group consisted of 26 patients, 6 of them still awaiting transplantation.

2.2 Questionnaire

It is generally recognized that health-related quality of life (or "health status") constitutes a complex, multi-dimensional construct (26). According to present standards objective and subjective components are discerned (27). Objective quality of life usually refers to observable phenomena which can be compared with external standards (e.g. walking distance). Subjective quality of life refers to experienced well-being. We selected a set of general and specific questionnaires which addressed objective and subjective of quality of life. If possible, validated Dutch questionnaires were used. The left-hand columns of table 3 show the questionnaires used as well as their ranges of scores and reference scores for the general population, if available.
The Nottingham Health Profile (NHP, part 1) is a comprehensive measure designed to measure perceived health on 6 specific domains of life, as shown in table 3 (28). The NHP consists of 38 items with a yes/no answering format and has been used by Lowe in his assessment of post-transplant status of liver transplantation patients (21). The Karnofsky-index is a global one-item measure for health status, often used in oncologic research (29). It covers domains like intensity of treatment and ability to take care of oneself. The Index of Well-being is a global measure for experienced well-being (24), consisting of 11 items. The other questionnaires mentioned in table 3 concern more specific indicators. The State-Trait Anxiety Inventory (STAI) and the Self-rating Depression Scale (SDS or Zung), are 20-item questionnaires to measure anxiety and depression respectively (30-33). The questions on Activities of Daily Life were derived from a Dutch national survey on health-related problems. For nine activities, ranging from dressing to shopping, patients were asked whether they performed these activities independently; and if they did, at what effort. The questions on physical complaints and working capacity were designed for this study. Inquiries were made about the following complaints by means of a three-point scale (absent/sometimes

present/always present): lack of appetite, abdominal cramps, swollen belly, itching, jaundice, bone pain, backache, haematomas, drowsiness. The questions on satisfaction with aspects of life originate from the Dutch health survey mentioned above.
The resulting questionnaire consisted of ± 250 items.
Stand-alone computer assisted interviewing was used as method of presentation of the questions and registration of the response (34). This technique was succesfully applied earlier with ambulatory as well as bedridden patients in a similar study of heart transplantation patients (35).
In addition to the self-reported quality of life, medical records of all patients were abstracted for the presence of psychiatric events. A psychiatric event was defined as clinical or outpatient treatment by a psychiatrist and/or the prescription of psychiatric drugs (excluding temporary prescription of benzodiazepine-derivatives).

3 Results

3.1 Response

Of 57 theoretically possible measurements among the 26 cross-sectional patients, 42 (75 %) were actually realized. From all of them at least one measurement was obtained. Fourteen measurements were missed (including 2 refusals) from 12 patients due to initial organizational reasons unrelated to the physical condition of the patient.
In the longitudinal group 22 of 26 possible pretransplant measurements (85 %) were obtained. Impaired physical condition of three patients prevented them from participation. From the 20 patients meanwhile transplanted, five died. From the survivors all 24 possible post-transplant measurements were obtained.
A total number of 88 (42+22+24) measurements was obtained from 46 patients (overall response rate: 82 %). The number of measurements related to the time of completion is presented in Table 1. As for paired observations (i.e. measurements obtained from the same patients before and after liver transplantation), 14 pre - 3 months pairs could be obtained.

Table 1 Number of questionnaires completed

pre	3m	1yr	2yr	3yr	4yr	5yr	6yr	7yr	8yr	9yr	10yr	Total
22	18	13	8	5	3	4	5	4	4	1	1	88

3.2 Patient characteristics

Socio-demographic and medical characteristics at the time of transplantation are summarized in Table 2. Only one of the interviewed patients showed clear signs of hepatic encephalopathy preceding liver transplantation. She completed half of the questionnaire with considerable assistance. There was no significant difference in severity of disease, diagnosis, or other patient characteristics between earlier or later transplanted patients in the study-group; liver function of those alive at 1 year follow-up or more was good (data not shown).

3.3 Quality of life

The interview results are summarized in Tables 3 and 4 and Figure 1.
The results shown in Table 3 represent a cross-sectional analysis. This implies that average values of measurements relate to groups of partially different composition. For reasons of presentation the results for the cross-sectional group of patients from 2 years after liver transplantation onwards were combined into two groups, resp. 2 to 5 years after liver transplantation and 6 years or more after liver transplantation (including only one measurement per patient per follow-up group). The results of the longitudinal analysis are shown in Table 4.
Finally we addressed the question whether physical complaints were related to restrictions on particular domains of quality of life. The severity of the self-reported complaints was correlated with the dimension scores of the Nottingham Health Profile and with the Index of Well-being. Only correlations (r) exceeding 0.55 are mentioned.

Table 2 Socio-demographic characteristics of the study population (n=46)

	n	%
Gender		
female	31	67
male	15	33
Educational level		
minimal education	10	22
intermediate education	30	65
high education	6	13
Diagnosis		
cirrhosis excl. PBC	12	8
primary biliary cirrhosis	17	37
prim. sclerosing cholangitis	6	13
re-transplantation	2	4
other	8	17
*Child-Pugh classification at transplantation**		
A	12	26
B	24	52
C	10	22
Child-Pugh score		
X (SD)	8.1	(2.1)
*Age (years) at liver transplantation**		
X (SD)	42.6	(11.1)

*minus 5 patients not yet transplanted

Table 3 Quality of life before and after liver transplantation (reference values, mean patient values and standard deviations; 1987-1989, n=46, cross-sectional analysis)

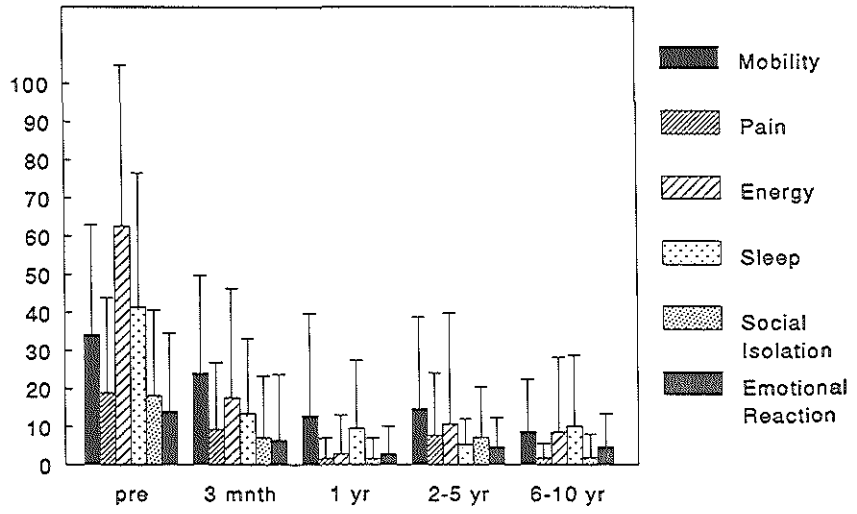
	Reference values		Results in liver transplantation-population				
Questionnaire	Range**	General Population	Waiting List (n=22)	3 Months post LTx (n=18)	1 Year post LTx (n=13)	2-5 Yrs post LTx (n=15)	6-10 Yrs post LTx (n=10)
General indicators							
Nottingham Health Profile - 1	100 - 0	≤ 15	***	***	***	***	***
Karnofsky-Index	0 - 100	≥ 90	64 (18)	71 (15)	87 (13)	85 (16)	9.6 (5)
Index of Well-Being	2.1 - 14.7	> 12	9.5 (2.9)	13.5 (1.1)	13.2 (2.0)	13.4 (2.3)	13.4 (2.0)
Specific indicators							
State-Trait Anxiety Inventory	80 - 20	≤ 37	41 (11)	34 (10)	29 (7)	30 (7)	3.2 (5)
Self-rating Depression Scale (Zung)	100 - 25	≤ 33	50 (11)	43 (8)	39 (5)	44 (7)	43 (6)
Activities of daily life	1 - 10	≥ 9	8.7 (1.7)	9.2 (1.4)	9.3 (1.9)	9.4 (1.4)	9.8 (0.6)
Physical complaints	10 - 1	≤ 2	5.6 (2.0)	3.4 (1.9)	2.6 (1.9)	3.4 (1.8)	2.8 (1.9)
Working* activity (hours/day)	0 - 12	± 8	1.9 (3.0)	1.7 (1.8)	3.3 (2.3)	4.1 (2.6)	5.0 (2.5)
Median value			0.0	1.0	4.0	4.0	5.0
Satisfaction with							
..health	5 - 1	< 2.6	4.1 (1.0)	2.2 (1.1)	1.6 (0.8)	1.6 (1.0)	1.4 (0.7)
..leisure time	5 - 1	< 2.6	3.2 (1.6)	2.3 (1.3)	1.7 (0.8)	1.7 (0.9)	1.5 (0.7)
..daily activities	5 - 1	< 2.6	3.1 (1.5)	2.3 (1.0)	1.7 (0.9)	1.8 (1.1)	2.0 (1.0)
..life as a whole	5 - 1	< 2.6	3.0 (1.1)	2.2 (1.2)	1.5 (0.7)	1.5 (0.8)	1.4 (0.5)

* paid and unpaid work (e.g. housework, study)
** worst possible score on the left, best possible score on the right
*** see figure 1

Table 4 Quality of life before and 3 months after liver transplantation (n=14, longitudinal analysis, t-test)

Questionnaire	Waiting List	3 Months after LTx	p-value
General indicators			
Nottingham Health Profile - 1			
Mobility	42	27	> 0.1
Pain	24	12	> 0.1
Energy	73	19	0.00
Sleep	47	16	0.01
Social Isolation	17	8	> 0.1
Emotional Reaction	23	8	0.07
Karnofsky-Index	57	71	0.02
Index of Well-Being	8.9	13.3	0.00
Specific indicators			
State-Trait Anxiety Inventory	39	34	> 0.1
Self-rating Depression Scale	51	43	0.04
Activities of daily life	8.4	9.1	> 0.1
Physical complaints	6.0	3.4	0.01
Satisfaction with			
..health	4.4	2.3	0.00
..leisure time	3.7	2.4	0.02
..daily activities	3.4	2.3	0.04
..life as a whole	3.4	2.1	0.01

Figure 1 Nottingham Health Profile part 1: mean scores before and after liver transplantation. 1987-1989.



mob = NHP-1 dimension "mobility"
pain = NHP-1 dimension "pain"
ener = NHP-1 dimension "energy"
sleep = NHP-1 dimension "sleep"
soc = NHP-1 dimension "social isolation"
emot = NHP-1 dimension "emotional reaction"

Restrictions on the Mobility-dimension correlated most with the presence of bone pain (r = 0.57), on the Pain-dimension with the presence of bone pain (r = 0.61) and backache (r = 0.67), on the Energy-dimension with a swollen belly (r = 0.60), on the Sleep-dimension with itching (r = 0.57), on the Emotional reaction-dimension with drowsiness (r = 0.59). None of the complaints correlated sufficiently with the score on the Social isolation-dimension. A low Index of Well-being correlated most with the presence of drowsiness (r = 0.59).
Only three patients needed psychiatric treatment. One patient underwent psychiatric consultation preoperatively because of a suicide attempt in the past. One patient needed anti-psychotic medication during the postoperative period. The third patient, who was actually transplanted for primary biliary cirrhosis, had a history of alcohol abuse. As anxiety to move hampered her remobilization, she was prescribed anxiolytic medication postoperatively.

4. Conclusion and discussion

In this article preliminary results of a study assessing the influence of liver transplantation on quality of life are shown. The pretransplant state can be characterized by a rather low Karnofsky index, psychological distress, many physical disturbances and a low level of experienced well-being. Three months after transplantation patients show a considerable rise of the quality of life level. Further improvement in the first post-operative year results in a quality of live level similar to or slightly below the level fo the general population. This level appears to stabilize in the following years. Obviously the low frequency of long term complications adds to this favourable picture. Experienced well-being shows an impressive favourable change following liver transplantation, the level exceeding that of the general population. The posttransplant NHP-scores resemble those reported by Lowe (20) closely, though the comparison should be made with caution as the latter study combines data from patients shortly after transplantation (43% < 1 year) with data of patients at more than 2 years follow-up.
We are aware that our conclusions depend on some assumptions especially with regard to the absence of a control group of nontransplanted patients and the effect of selective non-response.

In this study, liver transplantation patients were used as their own controls. As gradual deterioriation of quality of life in the nontransplantation case is likely (36), a comparison of quality of life-values before and after liver transplantation may give a conservative estimate of the positive effect of liver transplantation.
Selective (non-)response is another threat to validity. High severity of disease precluded three patients, who can be expected to benefit most from liver transplantation, to complete the pretransplant questionnaire. Measurements from 5 posttransplant patients were missed as a result of mortality. The overall result of this selective non-response might be a slight overestimation of average pre- and posttransplant quality of life.
Some results deserve attention. Firstly, the fairly good activities-of-daily-life score preceding liver transplantation seems to contrast with the limited scores on work-status and Karnofsky-index. Liver transplantation candidates are able to perform independently most daily activities only with great effort. Following liver transplantation these activities can be performed with little if any effort.
Secondly, the scores after liver transplantation on the SDS-Zung scale for depression seem rather high, i.e. in the range of Dutch non-depressive psychiatric patients (34,37). As reported by Smith, a high score does not necessarily mean the presence of a clinical depression following DSM-III standards (38). Although the SDS-Zung scale is an accepted and validated questionnaire for depression, its interpretation for somatic patients remains to be established.
Thirdly, psychiatric morbidity occurred apparently only infrequently. The support by a specialized social worker probably contributes to this favourable outcome. Surman reported frequent episodes of pre- and postoperative anxiety and depressive disorders, and considered psychiatric consultation an essential support to the transplant programme (17).
These results strongly suggest that liver transplantation not only improves survival, but also significantly improvements in quality of life. These findings should be validated by a longitudinal study, for which data are currently collected now.

5. Acknowledgement

The authors wish to thank Mrs. C. Steenmeyer, Mrs. M. Messchendorp and Mrs. E.B. Haagsma for their help.

References

1. Neuberger JM, Adams DH. Liver transplantation. Bailliere's Clinical Gastroenterology 1989; 3: 231.

2. Bonsel GJ, Klompmaker IJ, Veer F van't, Habbema JDF, Slooff MJH. Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis. Lancet 1990; i: 493.

3. Bonsel GJ, Klompmaker IJ, Essink-Bot ML, Habbema JDF, Slooff MJH. Cost-effectiveness analysis of the Dutch liver transplantation programme. Transplant Proc 1990; 22: 1481.

4. Starzl TE, Demeter AJ, Thiel D van. Liver transplantation (first of two parts). N Eng J Med 1989; 321: 1014.

5. U.S. Department of health and human services. Assessment of liver transplantation. Health Technology Assessment Reports, number I. Rockville, MD: Agency for Health Care Policy and Research, Public Health Service, Department of Health and Human Services, 1990.

6. Pennington JC. Quality of life following liver transplantation. Transplant Proc 1989; 21: 3514.

7. Starzl TE, Koep LJ, Schroter GPJ, Hood J, Halgrimson CG, Porter KA. The quality of life after liver transplantation. Transplant Proc 1979; 11: 252.

8. Rolles K, Williams R, Neuberger J, Calne R. The Cambridge and King's college hospital experience of liver transplantation, 1968-1983. Hepatology 1984; 4: 50S.

9. Scharschmidt BF. Human liver transplantation: Analysis of data on 540 patients from four centers. Hepatology 1984; 4: 95S.

10. Williams JW, Vera S, Evans LS. Socioeconomic aspects of hepatic transplantation. Am J Gastroenterology 1987; 82: 1115.

11. Krom RAF, Wiesner RH, Rettke SR, et al. The first 100 liver transplantations at the Mayo Clinic. Mayo Clin Proc 1989; 64: 84.

12. Eid A, Steffen R, Porayko MK, et al. Beyond 1 year after liver transplantation. Mayo Clin Proc 1989; 64: 446.

13. Gartner JC, Zitelli BJ, Malatack JJ, Shaw BW, Iwatsuki S, Starzl TE. Orthotopic liver transplantation in children: two-year experience with 47 patients. Pediatrics 1984; 74: 140.

14. House R, Dubovsky SL, Penn I. Psychiatric aspects of hepatic transplantation. Transplantation 1983; 36: 146.

15. Tarter RE, Thiel DH van, Hegedus AM, Schade RR, Gavaler JS, Starzl TE. Neuropsychiatric status after liver transplantation. J Lab Clin Med 1984; 103: 776.

16. Surman OS, Dienstag JL, Cosimi AB, Chauncey S, Russell PS. Psychosomatic aspects of liver transplantation. Psychother Psychosom 1987; 48: 26.

17. Foley TC, Davis CP, Conway PA. Liver transplant recipients - self-report of symptom frequency, symptom distress, quality of life. Transplant Proc 1989; 21: 2417.

18. Wolcott D, Norquist G, Busuttil R. Cognitive function and quality of life in adult liver transplant recipients. Transplant Proc 1989;21:3536.

19. Roberts MS. Quality-of-life measures in liver transplantation. In: Quality of life and technology assessment. F. Mosteller and J. Falotico-Taylor, ed. Washington DC: National Academy Press. 1989.

20. Lowe D, O'Grady JG, McEwen J, Williams R. Quality of life following liver transplantation: a Meliminary report. J Roy Coll Pract. 1990; 24: 43-46.

21. Tarter RE, Erb S, Biller PA, Switala JA, Thiel DH van. The quality of life following liver transplantation: a preliminary report. Gastroenterol Clin North Am 1988; 17: 207.

22. A technology assessment of liver transplantation: a study of the Dutch liver transplantation programma in Groningen 1977-1987 (in Dutch). Bonsel GJ, Habbema JDF, Bot ML, Veer F van't. Charro FT de, Maas PJ van der. Ned Tijdschr Geneeskd 1989; 133: 1406.

23. Lohr KN. Conceptual background and issues in quality of life. In: Quality of life and technology assessment. F. Mosteller and J. Falotico-Taylor, ed. Washington DC: National Academy Press, 1989.

24. Campbell A, Converse PE, Rodgers WL. The quality of American life. New York: Russel Sage Foundation, 1976. ISBN 87154 194 7

25. Putten ABMM van der, Bijleveld CMA, Slooff MJH, et al. Selection criteria and decisions in 375 patients with liver disease, considered for liver transplantation during 1977-1985. Liver 1987; 7: 84.

26. Slooff MJH, Bams JL, Sluiter WJ, et al. A modified cannulation technique for veno-venous bypass during orthotopic liver transplantation. Transpl Proc 1989; 21: 2328.

27. Klompmaker IJ, Haagsma EB, Gouw ASH, et al. Azathioprine and prednisolone immunosuppression versus maintenance triple therapy including cyclosporine A for orthotopic liver transplantation. Transplantation 1989; 48: 814.

28. Hunt SM, McEwen J, McKenna SP. Measuring health status. London: Croom Helm: 1986.

29. Grieco A, Long CJ. Investigation of the Karnofsky performance status as a measure of quality of life. Health Psychol 1984; 3: 129.

30. Spielberger CD, Gorsuch RL, Lushene RE (eds). STAI manual for the state-trait anxiety inventory. Palo Alto, California: Consulting Psychologists Press Inc., 1970.

31. Ploeg HM van der. Validation of the state - trait anxiety inventory, Dutch version (in Dutch). Tijdschr Psychol 1980; 35: 243.

32. Zitman FG, Griez EJL, Hooijer Chr. Standardization of depression questionnaires (in Dutch). Tijdschr Psychiatr 1989; 31: 114.

33. Zung WWK, Durham NC. A self-rating depression scale. Archiv Gen Psych 1965; 13: 63.

34. Erdman HP, Klein MH, Griest JH. Direct patient computer interviewing. J Consult Clin Psychol 1985; 53: 760.

35. Bonsel GJ, Bot ML, Boterblom A, Veer F van't. The costs and effects of heart transplantation. Volume 2C: Quality of life before and after heart transplantation, results (in Dutch). Department of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands, 1988. ISBN 90 72245 15 6

36. Christensen E. Individual therapy-dependent prognosis based on data from controlled clinical trials in chronic liver disease. Dan Med Bull 1988; 35: 167.

37. Dijkstra P. The self-rating depression scale of Zung (in Dutch). In: Stemming en ontstemming. Praag HM van, Rooymans HGM, eds. Amsterdam: De Erven Bohn BV; 1974.

38. Smith MD, Hong BA, Robson AM. Diagnosis of depression in patients with end-stage renal disease. Am J Med 1985; 79: 160.

Utility measurement of health states by the general public using computer assisted interviewing

Gouke J. Bonsel
Ben A. van Hout
Martin J. Verwey
J. Dik F. Habbema
Wim E. Saris
Paul J. van der Maas
Frank T. de Charro

Manuscript 1987

Abstract

A computer assisted interview schedule was developed for the valuation of health states. After minor modification of standard techniques, the interview was used to elicit value judgments of a laymen panel on a large number of different health states. This article presents the results of two subsequent investigations. The first study adressed feasibility and reliability of the procedure. Health state descriptions turned out to be too long and some valuation methods obviously were too difficult. In the second study a simplified version of the questionnaire was used. Forty health states were evaluated in interview sessions taking on average 25 minutes. Judgments consisted of scores on a fixed interval scale, which appeared to be reliable and internally valid on an individual level. Moreover, high consensus was found between individuals with respect to the ranking and the valuation of the health states. The application of the results in cost-utility analysis is discussed. We conclude that computer assisted interviewing may be a useful method to collect value judgments of the general public on different health outcomes. In this setting, line production turns out to be the valuation method of choice.

1. Introduction

Governments and health insurance companies increasingly have to decide on the incorporation of new medical technologies into existing insurance schemes. Cost-benefit analysis, cost-effectiveness analysis and cost-utility analysis are methods which can be of aid to the required decision-process. The authors are involved in the cost-utility analysis of the Dutch heart and liver transplantation programmes[1,2] These studies formed the impetus for the studies described in this paper.

Cost-utility analysis is regarded as a special form of cost-benefit analysis, in which benefits are represented by the change of life expectancy, weighted for the level of quality of life[3]. These benefits are referred to as QALY's (quality adjusted life years). The construction of QALY's was proposed by Fanshel and Bush first and implies the weighting of years of survival by means of an index representing the quality of the years survived[4]. The quality-index should be a cardinal value ranging from 0 for the worst possible health state to 1 for the best health state[5]. The determination of the quality-index in the context of CUA implies a valuation process

involving many theoretical and practical choices. Theoretical considerations for the construction and application of QALY's are given by e.g. Pliskin[6,7,8,9,10]. Some examples exist of the application of empirically based QALY's[11,12,13].
This paper presents the results of two subsequent studies which aimed at the determination of the cardinal values of health states using a computer assisted interview. Two questions were to be answered:

- is there a feasible, reliable and valid way of determining these values given the use of computer assisted interviewing?
- can the results be applied in cost-utility analysis?

In the second section some of the methodological concepts involved are discussed, followed by the description of the interview design and the study population chosen. Next the results of a pilot study (referred to as 'Study I) are presented, particularly on feasibility and reliability of the measurement instrument. The final sections present the results of the main study (Study II) including data on the degree of consensus between the respondents. We finish with a general discussion on the applicability of methods and results of these studies.

2. Methods

2.1 General approach

Direct valuation of health states appears to be a complex procedure from a methodological point of view. This valuation procedure essentially consists of the following steps:

1. the conception of an operational definition of health status and the estimation of the impact of a health programme on health outcome according to this definition (*health state discription*)
2. the construction of health scenarios and estimation of the economic value of these health scenarios (*health state valuation*)
3. the application of the results in cost-utility analysis (*implementation*)

Each of these steps involves many choices for which literature offers different options (see Table 1 on next page for details on the first two steps). In most cases a 'perfect' choice is not possible due to the fact that optimization of conflicting properties of the measurement instruments is required.

Table 1 Methodological choices in the description and valuation of quality of life

Measurement of the patient's health state

1. concepts of a health outcome measure for utility analysis
   - QALY
   - standard gamble utility
2. dimensions of health
   - physical
   - medical (symptoms)
   - psychological
   - future outlook
   - social relations
   - age
   - sex
   - work status
   - marital status
   - duration (time)
3. measurement level within a health dimension
   - nominal
   - ordinal
   - interval
   - ratio
4. data collection
   - observors
   - patient questionnaire

Valuation of health states or 'scenarios'

5. relation of patient's health state measurement with description in valuation procedure
   - implicit (expert opinion)
   - explicit (algorithm)
6. jury
   - doctors
   - nurses
   - patients
   - 'experienced' laymen
   - random sample of laymen
   - health services researchers
7. subject of health state description
   - 'I' (first person)
   - 'He' (third person)
8. presentation style
   - short hand
   - narrative style
9. valuation method
   - magnitude estimation
   - standard gamble
   - time trade-off
   - willingness to pay
10. interview situation
    - written interview with/without support
    - oral interview
    - computer assisted interview
11. sample technique

2.2 Health state description

The description and valuation of health states depend much on their underlying concept[14]. This concept is characterized by the dimensions which are distinguished. In spite of the growing literature on the measurement of health states, little information is available about the influence of health concepts and of the dimensioning

on reliability and validity of the results[15,16,17]. Additionally, the selection and operationalization of dimensions show much variety. Most authors use scaled categories within a dimension. The resulting structure is used to measure health outcome consequences at the patient level.

2.3 Health state valuation

From these observations we must derive formal health state 'summaries' or 'scenarios'. The same health concept and dimension-structure will usually be chosen, though simplification may be needed.
If the resulting scenarios have to be valued for economic purposes it should additionally be made clear whether a health scenario is to be considered by the judging subject as applying to him/herself - 'first person' - [14,18,19,20] or to someone else - 'third person' - [21,22,23,24] . Health scenarios can be written in a shorthand style [21,22,25] or in a narrative style, including or excluding statements on duration [14,20,26].
Regardless the method of presentation there exist many methods to elicit utility-judgments on these health scenarios[27,28,29,30,31,32,33,34]. The methods vary a.o. with respect to the resulting measurement level. In cost-utility analysis a cardinal scale is required to construct QALY's[33]. Consequently the rather easy method which ranks health scenarios on an ordinal scale (category rating) can not be used[35]. Below four other methods are considered: magnitude estimation, standard gamble, time trade-off and willingness to pay. More methods are described in the above mentioned reviews.
Magnitude estimation was originally developed by Stevens in 1958[36,37]. Respondents express their judgment about a health state directly in numbers. The assumed scale usually has two fixed endpoints, one end representing the least preferable health state - usually death - and the other end representing optimal health. This procedure yields a relative scale. Its usefulness in measuring subjective judgments has been demonstrated (see e.g. Bohrnstedt[38]) and its application in medicine is not limited to health state valuation[39]. Although the instructions of a magnitude estimation scale aim at a ratio scale, the obtained scores may have a loginterval level[40,41]. Other response modalities than numbers are possible e.g. visual analogue scales.
Standard gamble also yields cardinal values. Essentially the respondent has to state which probability p of immediate death combined with probability (1-p) of life-long optimal health is equivalent to the certainty (p=1) of a specific non-optimal health

state with 'infinite' (life-long) duration[42]. This method cannot easily be applied, since the cognitive task of comparing probabilities and different outcomes appears to be difficult, especially when small probabilities are involved [14,43]. Sometimes a visual aid for standard gamble is applied, bridging the difference between magnitude estimation and standard gamble. In addition standard gamble does not satisfy a constraint imposed by the construction of QALY's, i.e. not to include a time dimension. Therefore we did not use the standard gamble method. *Time trade-off* was developed specifically for the use in health care. Like standard gamble it is an indirect technique as it derives the preference for a particular health state from the respondent's response to a decision situation. The respondent is offered two alternatives. Alternative 1 implies being in health state A for a defined period (t_1) as opposed to alternative 2 which implies optimal health for a shorter time (t_A; $t_A < t_1$). Time t_A is varied until indifference between alternative 1 and 2 is reached. The preference value for A is easily calculated ($v_A = t_A/t_1 * v_1$; v_1 being the value for optimal health, which is usually set 1). The time trade-off method fits well to the constraint on the quality index in CUA with respect to the time dimension, which is an advantage as compared to standard gamble. Like standard gamble time trade-off involves a rather difficult process of determining respondent's indifference between alternatives with different probabilities.

Willingness to pay can be regarded as another trade-off method. Essentially the respondent is asked how much he would be willing to pay to get a well-defined gain in health[44]. Although from an economic point of view several authors have claimed this method is superior, its position is yet not well established[30,44,45,46].

As compared to the time trade-off method, the heterogeneous attitude of repondents towards money may be a disadvantage. Money is not a very good yardstick if respondents put very different values on money[45,46].

The choice of the appropriate *jury* or *panel* to obtain the values from, is influenced by the ultimate goal of the valuation study. Judgment is sometimes by expert panels consisting of doctors or nurses[25,47]. If results are implemented in a CUA, health states should be valuated from a societal point of view, suggesting a sample of the population to be the best panel, as in our study[18,25,33].

2.4 Criteria for evaluation

We regarded the following criteria to be important: feasibility, reliability and validity[48,49]. A measurement procedure is considered feasible if it is acceptable to the respondents, if it is easy to interviewers and to data-analysts, if it enables to measure a substantial number of health states, and if it can be performed at low costs. Reliability is related to the question to which extent the same phenomenon can be measured a second time with the same results. Reliability depends partially on acceptability to the respondents. A measurement procedure is valid if actually measures the phenomenon it claims to measure. This presumes a comprehensive health state description, including all relevant aspects. Consensus between respondents and the possibility of assuming a ratio scale can be regarded as specific aspects of validity in the context.

2.5 Interview design in Study I & II, and study population

Usually a structured and strongly supported interview situation is chosen for the valuation task. This approach involves lengthy (1.5 to 4.5 hours) interviews[50,51]. Although this approach to a certain extent guarantees adherence to the judgment task, judgments may be biased by the influence of supporting interviewers and by fatigue of the respondents[50]. A related disadvantage is the limited number of scenarios which can be valued this way. In our studies health states were presented and valued by means of a personal computer (computer assisted interviewing, CAI)[52]. Advantages of this method are the short response time, the option of incorporating special features (e.g. individual random presentation of questions, conditional presentation of questions) and generally good quality of data. All response modalities are possible e.g. categories, numbers and lines. Automatically the time required for each separate judgment task is recorded. If an elaborated 'help menu' is provided for, an unbiased supported interview situation can be approximated.
In both studies the same panel of respondents was involved, which consisted of a sample of households in the city of Amsterdam.
Each adult member of the household voluntarily participates in answering interviews using CAI on a monthly base. As a reward the panel members are given a home-computer in loan and occasionally some new software.

The home computer is loaded (survey questions) and unloaded (respondent's answers) automatically using the telephone communication line. Information on background variables is described in Table 2.

Table 2 Age, sex, education and income of study population (income in Dutch florins 1 Hfl ≈ 0.3 £) n=72

	n	%		n	%
Age			**Education**		
16 - 20 years	15	21	Primary school	4	6
20 - 30 years	13	18	Lower vocational	14	19
30 - 40 years	15	21	Secondary school	16	22
40 - 50 years	12	17	Middle vocational	5	7
50 - 60 years	9	16	College	13	18
60 - 70 years	7	10	Higher vocational	15	21
> 70 years	1	1	University	5	7
Sex			**Income if employed**		
Male	35	49	<1000	5	15
Female	37	51	1000 - 1500	4	12
			1500 - 2000	4	12
			2000 - 2500	8	24
			2500 - 3000	5	15
			3000 - 3500	3	9
			>3500	4	12

3. Study I

3.1 Design - additional features

In Study I 21 different health scenario's with 6 dimensions were constructed. The health concept was partially based on the early work of Patrick and Bush[21,22]. The following dimensions were used (in order of presentation): physical functioning, medical symptoms, social relationships, psychological well-being, time spent in the health state and intensity of treatment. The psychological dimension was included according to the W.H.O.-definition of health[53,30], and to the recent literature on quality of life measurement (see e.g. De Haes and Katz[54,55]). The intensity of the medical treatment was added because in the post-transplantation phase there is often a striking discrepancy between apparent health state and the need for medical treatment. Finally a statement about the duration of the health state was incorporated. No information was included about the name of the particular disease or medical condition. A disease label introduces in fact extra information which might influence respondents in an uncontrolled way[18].
Other information e.g. on the patient's social status and work status was not included. Since we restrict ourselves to the question whether different health states can be valuated as a contribution to rational health care resource allocation decisions, only aspects of the health situation itself seem relevant. If distributional effects of resource allocation decisions should be taken into account too, other aspects of the health situation may be included, e.g. age.
In a modified factorial design 21 health states were defined, after exclusion of some too unrealistic health states. The interview started with an introduction on the topic and with a presentation of the standards i.e. the best and the worst state within our classification. Next the respondent valued the health states in blocks, each block comprising of one valuation method. Only a selection of health states was evaluated by the time trade off and willingness to pay method.

3.2 Feasibility

In this case non-response is not a good indicator of acceptability of the judgment task as members of this panel could be expected to accept the task even if they consider the task too difficult. Four other indicators of the feasability seem more

appropriate: incomplete response, answer patterns, comments of respondents and interviewing time.
Few respondents broke off the interview, but on the answer patterns both trade-off tasks turned out to be too complex. About half of the respondents clearly gave nonsensical responses and 20% refused to answer. The small number of interpretable answers appeared to be inconsistent as nearly identical responses were given for different stimuli, while different responses were given on very similar stimulus.
Respondents were given the opportunity to comment at the end of the interview. In study I most respondents commented on the interview. Descriptions were considered to be too long and too similar; particularly the willingness to pay method was regarded to be unethical, notwithstanding a comprehensive introduction preceding this valuation method. The trade-off tasks (time trade-off and willingness to pay) prolonged the interview unacceptably. The whole interview in Study I was completed in about 45 minutes. The trade-off tasks accounted on average for 20 minutes, although they constituted 19% of the questions.
With regard to other aspects of feasibility (e.g. interviewers-acceptability, flexibility to change, data quality, cost) the interview method appeared to be satisfactory.

3.3 Reliability

Reliability can be estimated by repeated measurement provided that the design guarantees that correlation cannot be due to memory effects and variation cannot be due to change in time of the phenomenon[52]. Memory effects seemed not likely to have occurred in our studies. Scenarios were randomly presented and respondents revealed concern about consistency when they commented on the interview. Standard test-retest reliability, repeating some questions with the same measurement method, was satisfactory. In addition a second test-retest procedure was applied, based on the assumption that magnitude estimation using numbers and magnitude estimation using line production are parallel instruments. Application of this type of test-retest procedure in Study I resulted in a median of correlation of only 0.67. This disappointing value was probably the result of too complex health scenarios.

3.4 Relation between dimensions and health state value

With linear regression we estimated the relation between the different dimensions and the total health scenario scores. For convenience log transformation was applied to the time dimension, which originally consisted of the categories '2 weeks', '6 weeks', '3 months', '1 year', etc. The other dimensions were scaled following their logical ordening, assigning 1 to the best category, 2 for the second best category, and so on. The assigned ordinal number of each category was conveniently regarded as a variate. Without interaction terms the model explained 88% of the total variance. The attribution of each dimension was fairly uneven, social relations and medical symptoms attributing insignificantly to the variance explained (data not shown).

4. Study II

4.1 Study design - additional features

The design of Study II was modified according to the results of Study I. Social relations as a dimension was excluded. The dimension time was also omitted for methodological reasons (QALY-construction). Finally some categories were reformulated.

Table 3 presents the operationalization of the four remaining dimensions, each consisting of five categories. Twenty-six health state descriptions were formulated by combining one category from each dimension. The combinations were partially based on empirical patient data from the heart transplantation study and partially constructed to achieve maximal variation.

The interview task started with valuation of all health states by line production. Next a selection of eleven health states was valued using number assignment. Within each block the order of presentation was random on the individual level. The trade-off tasks were omitted.

Table 3 Health state dimensions and categories of Study II (translated)[1]

A. Intensity of treatment
1. at home
2. at home, but having regular medical check-ups
3. at home, under intensive out-patient control, taking powerful drugs
4. temporarily hospitalized
5. in hospital, intensive treatment

B. Medical symptoms
1. no complaints or disorders
2. some small complaints, but not in pain
3. rather quickly short of breath, at times in pain
4. at the smallest effort short of breath and tired, in moderate pain
5. constantly in severe pain and out-of-breath

C. Physical functioning
1. able to do everything both at home and outside
2. able to do everything at home, but restricted outdoors (e.g. unable to cycle or do the shopping)
3. able to care for him/herself physically, though limited in walking around
4. difficulties with getting in/out of bed needing help with self-care
5. restricted to bed, unable to do anything

D. Psychological well-being and subjective medical prognosis
1. no psychological complaints, and having trust in the future
2. feeling down now and then, but having trust in the future
3. under stress having trust in the future, but without looking too far ahead
4. depressed moderate trust in the future, does not look far ahead
5. tense and depressed, living day by day, and fearing the worst

[1] One specific health scenario can be indicated by four digits, representing the category-number of dimensions A, B, C, and D respectively. Thus optimal health is health scenario 1111.

Table 4 presents the line judgments of the panel on 26 different health states. Health states are indicated by a 4 digit number, each digit representing the respective category of a dimension. The dimensions and categories are ordered as in Table 3. Scores are transformed from a 1 - 39 to a 0.0 - 1.0 scale.

Table 4 Line judgments of 26 health states. Health states ordered according to the arithmetic mean (n=65).

Health State	Aritm. Mean	S.D.	95% Conf. Interval	Geom. Mean	Median
1111	0.97	0.06	0.95-0.98	0.96	1.00
3111	0.60	0.19	0.55-0.64	0.56	0.59
2222	0.49	0.18	0.45-0.54	0.45	0.49
2232	0.47	0.20	0.42-0.52	0.42	0.46
1331	0.46	0.19	0.42-0.51	0.42	0.43
2322	0.43	0.16	0.40-0.47	0.40	0.41
4222	0.43	0.19	0.39-0.48	0.38	0.43
3114	0.40	0.19	0.35-0.44	0.34	0.41
1323	0.39	0.16	0.35-0.43	0.35	0.38
3321	0.36	0.15	0.32-0.40	0.32	0.32
4432	0.36	0.17	0.32-0.40	0.30	0.35
2224	0.36	0.16	0.32-0.39	0.31	0.38
3323	0.33	0.17	0.29-0.37	0.28	0.30
3333	0.32	0.15	0.29-0.36	0.28	0.30
5232	0.30	0.18	0.25-0.34	0.24	0.27
2442	0.29	0.16	0.25-0.33	0.23	0.27
3343	0.27	0.14	0.24-0.31	0.23	0.27
2434	0.27	0.15	0.23-0.31	0.22	0.27
4434	0.25	0.15	0.22-0.29	0.20	0.24
4444	0.20	0.13	0.17-0.23	0.15	0.19
4553	0.18	0.11	0.15-0.21	0.14	0.16
5443	0.15	0.11	0.12-0.17	0.11	0.11
4454	0.14	0.10	0.12-0.17	0.11	0.14
5543	0.13	0.11	0.11-0.16	0.09	0.08
4545	0.11	0.10	0.08-0.13	0.07	0.05
5555	0.01	0.02	0.01-0.01	0.01	0.00

The modifications of the interview improved the acceptability of the interview. Nonsense answers were absent and comments of the panel were predominantly positive. The average interviewing time, instruction time included, was reduced to 25 minutes. Shorter interviewing time should not be expected as this type of judgment is cognitively and emotionally difficult.

4.2 Reliability

With the same test-retest procedure as in Study I, regarding the different answer modes as parallel instruments, the median correlation rose to 0.92, which may be regarded as a very satisfactory result (see Table 5).

Table 5 Correlation between line judgments and number judgments on the individual level

Study I	median of correlation	.67
	mean of correlation	.66
Study II	median of correlation	.92
	interquartile-distance	.89 - .95
	mean of correlation	.89

4.3 Validity

Four types of validity can be distinguished according to the Recommendations of the APA Committee on Psychological Tests[56]: predictive validity, concurrent validity, content validity and construct validity, the first two taken together as criterion validity. Criterion validity would require the prediction of particular health behaviour or health care expenditures depending on stated values for different health states. Sofar, this type of validation seems to involve too complex methodological and empirical problems. In fact this is exactly the raison d'être of the second best methods described. Content validity seems not relevant in our case as health states are not regarded as psychological construct but as a commodity[30].

To our opinion in this study only construct validity can be determined. This type of validity is involved whenever a test is to be interpreted as a measure of some attribute or quality which is not itself 'operationally defined'[56]. Construct validity is obtained deductively and involves a process in which evidence from different sources is integrated. In this section we consider 5 different indications for construct validity: order effects, scale distribution, logical consistency, independency of dimensions and results from linear regression after rescaling the categories.

No effects due to the *order of dimensions* in the health scenario should exist. We varied the position of the time dimension in Study I. Some evidence in support of an internal order-effect was found, but no definite conclusion could be drawn due to the low reliability of Study I. One might expect that this phenomenon will occur when the list of dimensions is long resulting in predominance of the first dimension. We repeated this experiment in Study II. Two health scenarios were presented twice with different ordering of the dimension statements. As no order-effect was found it may be tentatively concluded that respondents attached a value to the health state as a whole.

Another criterion of construct validity is the *distribution of values* on the response-continuum[57], both individually and on the group level. Categorical answers were found in Study I due to apparent difficulties with the judgment task. No such response behaviour was found in Study II on the individual level. On the aggregate level mean values were equally distributed along the response continuum except for a 'gap' between optimal health (state 1111) with value 0.97 and the next-best health state (state 3111) with value 0.61. The same result was reported by Patrick[21] but Rosser and Kind showed a great majority possible health states was given a value between 0.9 and 1.0[50]. Torrance[3], using multi-dimensional scenario's like those presented in this article, reported mean values which were more equally distributed. Furthermore we obtained the mean judgments about different health states expressed both in lines and - independently - in numbers. When individual reliability is high the coefficient of identity between both group scales[58] can be computed as a measure of internal validity of such a group scale. This coefficient was equal to 0.98 and the aggregated group scales are thus exchangeable.

Logical/consistency provides us with a third validity criterion. If two health state scenarios A and B are compared and for all dimensions the respective category of A has a higher (lower) or equal rank compared to B, then the expressed value of A should be at least (most) equal the value of B, assuming no interaction of dimen-

sions. E.g. the value of state 1323 should at least equal the value of state 3323; no logical ordering exists between e.g. 1323 and 3114 and as a consequence no prediction is possible about the order of their respective values. Inconsistent response behaviour can be related to the 'distance' between two logical ordered states. We define 'distance' conveniently as the sum of category-rank differences of all (4) dimensions. This sum ranges from 1 (e.g. states 2222 and 2232) to 16 (states 1111 and 5555).
The 26 different health states accounted for 206 logically ranked pairs. Respondents expressed in 92 % of all logical pairs consistent values and expressed in 8% reversed preference. Relating inconsistent response behaviour to distance between a pair of health states showed a strong association: the majority (78%) of inconsistencies occuring when the distance was less than 5 (see Table 6).

Table 6 Consistency according to category - rank difference between pairs of health states

Distance	Pairs	Consistent order	Reversed order	
		Better	Equal	Worse
1	7	55.4%	14.7%	29.9%
2	13	69.1%	10.3%	20.6%
3	20	78.5%	6.9%	14.6%
4	33	81.1%	5.8%	13.1%
5	27	88.6%	3.9%	7.6%
6	17	88.4%	3.7%	7.9%
7	22	92.5%	2.7%	4.8%
8	25	95.4%	2.1%	2.5%
9	15	97.1%	1.3%	1.5%
10	7	97.1%	1.1%	1.8%
11	9	98.5%	0.9%	0.7%
12	5	99.7%	0.3%	0.0%
13	3	100.0%	0.0%	0.0%
14	2	100.0%	0.0%	0.0%
16	1	100.0%	0.0%	0.0%
Total	206	87.2%	4.3%	8.5%

The response behaviour turned out to be consistent, inconsistencies being related to resemblance of health states.
In the context of a multi-attribute utility function each dimension should prefarably *independently attribute* to the resulting overall health state value. If the attribution of dimensions does not show interaction, a change from e.g. category 1 to category 3 in one dimension should result in about the same difference between the scores, regardless the level of the other dimensions. Four pairs of health states could be selected which allowed such a check of validity (see Table 7).

Table 7 Interaction between dimensions

Dimension 1				
1 1 1 1	versus	3 1 1 1	0.37 }	difference 0.31
1 3 2 3	versus	3 3 2 3	0.06 }	
Dimension 1				
2 2 2 2	versus	4 2 2 2	0.06 }	difference 0.04
2 4 3 4	versus	4 4 3 4	0.02 }	
Dimension 3				
2 2 2 2	versus	2 2 3 2	0.31 }	difference 0.20
3 3 2 3	versus	3 3 3 3	0.11 }	
Dimension 4				
2 2 2 2	versus	2 2 2 4	0.14 }	difference 0.03
4 4 3 2	versus	4 4 3 4	0.10 }	

The first two rows of this table show that a change rise from category 1 to category 3 in dimension 1 shows a difference of the total score of 0.37 and 0.06 respectively. Only a minor difference should be expected as is e.g. the case in the second and the fourth pair. We conclude that the attribution of each dimension to the total score is not always independent from the level of the categories in the other dimensions. One might argue this interaction should be expected, assuming that simple additivity of the disutilities of suboptimal health aspects does not exist[59].
The relative *contribution of each dimension to the group scale* was estimated, applying ordinary lineair regression. The mean score of a health state was the dependent variable and the corresponding categories of each dimension, scored as a variate from 1 to 5, acted as the explanatory variables. This model explained 83 % of total variance. As the numeric scale from 1 to 5 is arbitrary, we investigated whether higher order models could improve the model. First we normalized the categories between 0 and 1. The best category was indexed 0 and the worst was indexed 1. The remaining three dimensions were supposed to be on a line:
$f(i) = i^{alfa}$, i=0.25, 0.5, 0.75
Table 8 presents the results of this rescaling procedure.

Table 8 Original and rescaled category values

Dimension	A	B	C	D
Category	Rescaled category value			
1	0	0	0	0
2	0.49	0.65	0.11	0.21
3	0.86	0.81	0.33	0.46
4	0.70	0.91	0.63	0.72
5	1	1	1	1

Note that this procedure lead to an category-order reversal (A3 - A4). Using the rescaled category-scores 91 % of total variance could be explained.

The estimated model is (A_i being the rescaled category-score on dimension A, etc. and T-values between parenthesis):

$$Y = 0.80 - 0.21*A_i - 0.28*B_j - 0.08*C_k - 0.20*D_l$$
$$\qquad\qquad (-4.2) \quad (-5.0) \quad (-1.4) \quad (-4.1)$$

The product moment correlation between the observed mean values of the 26 health states and the predicted values was 0.96, which is an acceptable value. This model enabled us to estimate the values of all possible health states within our classification.

4.4 Consensus

In this section we address the question whether a common scale of health state values does exist. Without the application of the resulting values for QALY-calculations in CUA, the issue of consensus would be of merely academic interest. However, implementation in CUA depends on consensus. The argument runs as follows. Suppose we have derived a period-dependent quality index, from the cardinal values created in the valuation process described before. I.e. the cardinal value of the quality-index is based on some measure of central tendency applied to the values of the respective health scenario(s) which apply to that period. Usually the aritmetic mean is chosen, but other measures of central tendency may be appropriate from a statistical point of view, for example the mode, median of geometric mean (see Table 4). Each measure of central tendency implies some rule for aggregation of preferences with an assumption about the measurement level. In the case of the aritmetic and geometric mean, this level is supposed to be cardinal, in the case of mode and median nominal and ordinal respectively.

It is important to be aware of the non-statistical assumptions of this aggregation. All measures of central tendency presuppose commensurability of the individual scales[60], which is a prerequisite for an empirically based social welfare function[61]. Assuming commensurability, the degree of dissensus should be defined according to the measurement level involved (standard deviation, inter-quartile-distance). If consensus appears from the data, application of the group values seems justified. However, dissensus causes serious problems. Suppose a health care program for the treatment of disease A is evaluated. The CUA yields the costs and benefits, the latter expressed in QALY's. Ultimately QALY's are derived from the mean (or

median, etc.) of value judgments of the health scenario(s) before and after treatment. For a future group of consumers of the treatment, or for society as a whole, these group values do not provide us with useful information if dissensus exist. Some will regard the cost per benefit, which results from the judging panel on the health scenarios to be rather low and they will consequently benefit from a consumer surplus. Others will experience a consumer loss, as consumption will have a more or less obligatory character.

Each aggregation principle in fact implies some rule for counting these benefits and losses. It can be argued that if no common rule for counting benefits and losses exists, the QALY construction is not justified when dissensus exist. Examination of our data for the existence of a common scale presupposes:

- sufficient reliability and validity of the answers on the individual level
- the assumption of commensurability of the individual scales
- a satisfactory aggregation technique for the individual answers

Assuming these conditions to be sufficiently fulfilled and assuming a cardinal measurement level, we estimated the degree of consensus between individuals. The line-answers were used of 65 respondents for whom which sufficient demographic data were available. The arithmetic mean was chosen to compute values of the group-scales. First individual scales are compared to the group-scale. The effect of outliers is shown. Finally the existence of subgroups according to three background variables is investigated.

The *correlation* between the individual scales and the overall group scale was estimated. For this association to be high, both a high intra-reliability and a high consensus are required. In Study II the mean correlation between the individual line scores and the group scale was 0.86. Presumably laymen agree to a large extent about the degree of severity of different health states. Additionaly we investigated the hypothesis whether respondents' individual scale - group scale correlation depended on three background variables: age, education and a variable representing the respondents' experience with health problems or health care. We did not find any significant association (Kendall's tau b insignificant in all three cases). Although these results seem sufficient on aggregate level, they do not exclude the possibility of substantial individual variation. The observed standard deviations of the valuations were about 0.16 on the 0.0 - 1.0 scale (see Table 4). This might be due (1) to respondents with a low intra-respondent reliability, (2) to respondents with a different opinion compared to the group-scale or (3) to

variation in response behaviour. Excluding the scores of the 13 respondents with the lowest intra-respondent reliability, the scores of the 4 respondents with the greatest deviation of opinion and the scores of 4 respondents with variation in response behaviour, the mean standard deviation only decreased 0.03. We thus concluded that no group of deviators could indicated this way.
Because of conflicting reports in literature, which sometimes showed characteristics of respondents influencing judgments[18,62] we tested in more detail the potential influence of background variables. The panel was dichotomized according to educational level, age and health care experience respectively. Subgroup scales were calculated and compared applying the coefficient of identity[58] - reflecting the degree to which two scales have the same mean, same dispersion and same distribution. No differences between subgroups could be demonstrated this way.

5. General discussion

Medical, psychological and economical science meet in the valuation of health states. The implementation of results in cost-utility analysis requires good technical qualities of the measurement instrument and unequivocal results.
After a pilot study we arrived at a four-dimensional health concept, each dimension consisting of five ordered categories. In the pilot study we could not replicate the results reported in other studies when we applied the willingness to pay and the time trade-off valuation technique. This may be caused by the fact that our panel was selected from the general public. Also the fact we applied a computer assisted interview, which limits the possibility of technical support of the respondent, has contributed to our selection of two magnitude estimation techniques. The results of Study II were satisfactory, showing reliable, consistent and homogeneous answers.
With regard to the requirement of a cardinal measurement level a definite answer is lacking, as no formal criterion exists with respect to this aspect of validity. Based upon comparison of different valuation methods (e.g. in Torrance[33]) one may assume a 'true' cardinal utility function which is a monotonic transformation of the empirically determined value function. Not knowing the exact relationship between our value function and the 'true' cardinal utility function, sensitivity analysis should be carried out if the results are applied in CUA[26,41].
Ultimately this study aimed at application of results in the medical technology

assessments of the heart and liver transplantation in the Netherlands. We derived health state descriptions (health scenarios) from survey-answers of patients who participated in the Dutch transplantation programmes. The earlier shown model was used to estimate health state utilities. Subsequently, sensitivity analysis was carried out as previous described. The difference between the mean of the estimated values of the health states before and after transplantation was 0.40 - 0.55. Thus also costs per QALY could be estimated[1,63].

6. Conclusion

We conclude computer assisted interviewing is a rewarding technique in valuation research. Despite some technical limitations, it turned out to be a feasible, reliable and valid measurement instrument. Application of the results in cost-utility analysis seemed justified. In this setting line production (linear analogue scaling) turned out to be the valuation method of choice. Certainly the next future will show the combination of audio-visual devices with a personal computer as interview instrument. This will enable the incorporation of other valuation techniques in the interview schedule which are indispensable for more information on the validity of the various valuation techniques.

7. Acknowledgements

The authors wish to thank the members of the European Common Core Group (Alan Williams, Paul Kind, and others) for their helpful comments.

References

1. Charro FT de, Bonsel GJ, Hout BA van. The costs and effects of hearttransplantation (in Dutch). Volume 9. Final Report. Rotterdam: Department of Economics (Faculty of Law) & Department of Public Health and Social Medicine, Erasmus University Rotterdam, 1988.

2. Habbema JDF, Bonsel GJ. The costs and effects of livertransplantation (in Dutch). Volume 1. Final Report. Rotterdam: Department of Public Health and Social Medicine, Erasmus University Rotterdam, 1988.

3. Torrance GW. Utility approach to measuring health-related quality of life. J Chronic Dis 40: 593-600, 1987.

4. Fanshel S, Bush JW. A health status index and its application to health-services outcomes. Operations Research 18: 1021-1066, 1970.

5. Haig THB, Scott DA, Wickett LI. The rational zero point for an illness index with ratio properties. Med Care 24: 113-122, 1986.

6. Pliskin JS, Beck CH. A health index for patient selection: a value function approach with application to chronic renal failure patients. Management Science 22: 1009-1021, 1976.

7. Torrance GW. Health status index models: a unified mathematical view. Management Science 22: 990-1001, 1976.

8. Pliskin JS, Shepard DS, Weinstein MC. Utility functions for life years and health status utility functions. Operations Res 28: 206-224, 1980.

9. Miyamoto JM, Eraker SA. Parameter estimates for a QALY-utility model. Med Decis Making 5: 191-213, 1985.

10. Ferguson BA. Working Towards a Cost Per QALY for Heart Transplantation. Dissertation submitted for M.Sc. in Health Economics. United Kingdom: University of York, 1985.

11. Bombardier C, Ware J, Russel IJ, Larson M, Chalmers A, Read JL. Auranofin therapy and quality of life in patients with rheumatoid arthritis. Results of a multicenter trial. Am J Med 81: 565-578, 1986.

12. Boyle MH, Torrance GW, Sinclair JC, Horwood SP. Economic evaluation of neonatal intensive care of very-low-birth-weight infants. N Engl J Med 308. 1330-1336, 1983.

13. Bush JW, Fanshel S, Chen MM. Analysis of a tuberculin testingprogram using a health status index. J Socio-Econ Plan Sci 6: 49-69, 1972.

14. Llewellyn-Thomas H, Sutherland J, Tibshirani R, Ciampi A, Till JE, Boyd NF. Describing health states. Methodologic issues in obtaining values for health states. Med Care 22: 543-552, 1984.

15. Quinn RJ. The effect of measurement method on preference scales in a medical decision-making context (Abstract). Med Decis Making 1: 431, 1981.

16. Drummond MF, Stoddart GL, Torrance GW. Methods for the economic evaluation of health care programmes. Oxford: Oxford University Press, 1987.

17. Loomes G. Disparities between health state measures: an explanation and some implications. Health Economics Study Group paper. York: Department of Economics University of York, 1988.

18. Sackett DL, Torrance GW. The utility of different health states as perceived by the general public. J Chron Dis, 31: 697-704, 1978.

19. Torrance GW. Multi-attribute utility theory as a method of measuring social preferences for health states in long-term care. In Values and long-term care. Kane RL and Kane RA (eds). California: Rand Corporation, DC Heath and Company, 1982.

20. Llewellyn-Thomas HA, Sutherland HJ, Tibshirani R, Ciampi A, Till JE, Boyd NF. The measurement of patients' values in medicine. Med Decis Making 2: 449-462, 1982.

21. Patrick DL, Bush JW, Chen MM. Methods for measuring levels of well-being for a health status index. Health Serv Res: 228-245, 1973.

22. Patrick DL, Bush JW, Chen MM. Toward an operational definition of health. J Health Soc Behav 14: 6-22, 1973.

23. Rosser R, Watts V. The measurement of illness. J Oper Res Soc 29: 529-540, 1978.

24. Pauker SG, Kassirer JP. Decision analysis. N Engl J Med 316: 251-258, 1987.

25. Kind P, Rosser R, Williams A. Valuation of quality of life: some psychometric evidence. In The value of life and safety. Proceedings of a conference held by the 'Geneva Association'. Jones-Lee MW (ed). Amsterdam: North-Holland Publishing Company, 1982.

26. Buxton M, Ashby J, O'Hanlon M. Alternative methods of valuing health states: A comparative analysis based on an empirical study using the time trade-off approach in relation to health states one year after treatment for breast cancer. Paper presented to the Third Annual Meeting of the International Society of Technology Assessment in Health Care, Rotterdam, 21-22 May 1987.

27. Kneppreth NP, Gustafson DH, Rose JH. Techniques for the assessment of worth. In Health status indexes. Berg RL (ed). Chicago: Hospital Research and Educational Trust, 1973.

28. Culyer AJ, Lavers RJ, Williams A. Social indicators: health. In Social Trends no. 2. Nissel M (ed) London: HMSO, 1971.

29. Culyer AJ. Measuring Health; lessons for Ontario. Toronto (Canada): University of Toronto Press. 1978.

30. Lipscomb J. Value preference for health: meaning, measurement, and use in program evaluation. In Values and long-term care. Kane RL and Kane RA (eds). California: Rand Corporation, DC Heath and Company, 1982.

31. McNeill BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. N Engl J Med 306: 1259-1262, 1982.

32. Read JL, Quinn RJ, Berwick DM, Fineberg HV, Weinstein MC. Preferences for health outcomes. Comparison of assessment methods. Med Decis Making 4: 315-329, 1984.

33. Torrance GW. Measurement of health utilities for economic appraisal. J Health Econ 5: 1-30, 1986.

34. Veit CT, Ware JE jr. Measuring health and health-care outcomes: issues and recommendations. In Values and long-term care. Kane RL and Kane RA (eds). California: Rand Corporation, DC Heath and Company, 1982.

35. Kaplan RM, Bush JW, Berry CC. Health status index. Category rating versus magnitude estimation for measuring levels of well-being. Med Care 17: 501-525, 1979.

36. Stevens SS. Psychophysics. New York: John Wiley, 1975.

37. Saris WE. Continuous scales in the social sciences: an attractive possibility. Amsterdam: Faculty of Political Sciences, University of Amsterdam, 1986.

38. Bohrnstedt GW. A quick method for determining the reliability and validity of multiple-item scales. Am Sociol Rev 34: 542-548, 1969.

39. Hsiao WC, Braun P, Yntema D, Becker ER. Estimating physician's work for a resource-based relative-value scale. New Eng J Med 319: 835-841, 1988.

40. Saris WE. Different questions, different variables. In Second generation of multivariate analysis. Fornell C (ed). New York: Praeger, 1982.

41. O'Brien BJ, Buxton MJ, Ferguson BA. Measuring the effectiveness of heart transplant programmes: quality of life data and their relation-ship to survival analysis. J Chronic Dis 40: 137S-153S, 1987.

42. Neumann J von, Morgenstern O. Theory of games and economic behavior. Princeton: Princeton University Press, 1944.

43. Tversky A, Kahneman D. Judgement under uncertainty: heuristics and biases. Science 185: 1124-1131, 1974.

44. Thompson MS, Read JL, Liang M. Feasibility of willingness-to-pay measurement in chronic arthritis. Med Decis Making 4: 195-215, 1984.

45. Thompson MS, Read JL, Liang M. Willingness-to-pay concepts for societal decisions in health. In Values and long-term care. Kane RL and Kane RA (eds). California: Rand Corporation, DC Heath and Company, 1982.

46. Mishan EJ. Cost-benefit analysis. London: Allen & Unwin, 1984.

47. Torrance GW, Sackett DL, Thomas WH. Utility maximization model for program evaluation: a demonstration application. In Health status indexes. Berg RL (ed). Chicago: Hospital Research and Educational Trust, 1973.

48. Torrance GW. Social preferences for health states: an empirical evaluation of three measurement techniques. J Socio-Econ Plan Sci 10: 129-136, 1976.

49. Kaplan RM, Bush JW, Berry CC. Health status: types of validity and the index of well-being. Health Serv Res: 478-507, 1976.

50. Rosser R, Kind P. A scale of valuations of states of illness: is there a social consensus? Int J Epidemiol 7: 347-358, 1978.

51. Buxton M, Ashby J, O'Hanlon M. Valuation of health states using the time trade-off approach: report of a pilot study relating to health states one year after treatment for breast cancer. London: Health Economics Research Group, Brunel University, Discussion Paper no. 2, 1986.

52. Saris WE, Pijper WM de. Computer assisted interviewing using home computers. European Research 14: 144-150, 1986.

53. Breslow L. A quantitative approach to the World Health Organization definition of health: physical, mental and social well-being. Int J Epidemiol 1: 347-355, 1972.

54. Haes JCJM de, Knippenberg FCE van. The quality of life of cancer patients: a review of the literature. Soc Sci Med 20: 809-817, 1985.

55. Katz S, et al. The Portugal conference: measuring quality of life and functional status in clinical and epidemiological research. J Chron Dis 40: 449-650, 1987 (special issue).

56. Cronbach LJ, Meehl PE. Construct validity in psychological tests. Psychol Bull 52: 281-302, 1955.

57. Drummond MF. Resource allocation decisions in health care: a role for quality of life assessments? J Chronic Dis 40: 605-616, 1987.

58. Zegers FE, Berge ten JMF. A family of association coefficients for metric scales. Psychometrika 50: 17-24, 1985.

59. Forst BE. Quantifying patient's preferences. In Health status indexes. Berg RL (ed). Chicago: Hospital Research and Educational Trust, 1973.

60. Hilden J. The nonexistence of interpersonal utility scales. A missing link in medical decision theory? Med Decis Making 5: 215-228, 1985.

EuroQol - a new facility for the measurement of health-related quality of life

The EuroQol Group

Health Policy, 16 (1990) 199-208
Published with permission of the editor

HPE 00371

EuroQol* – a new facility for the measurement of health-related quality of life

The EuroQol Group**

Accepted 14 August 1990

Summary

In the course of developing a standardised, non-disease-specific instrument for describing and valuing health states (based on the items in Table 1), the EuroQol Group (whose members are listed in the Appendix) conducted postal surveys in England, The Netherlands and Sweden which indicate a striking similarity in the relative valuations attached to 14 different health states (see Table 3). The data were collected using a visual analogue scale similar to a thermometer (see Table 2). The EuroQol instrument is intended to complement other quality-of-life measures and to facilitate the collection of a common data set for reference purposes. Others interested in participating in the extension of this work are invited to contact the EuroQol Group.

EuroQol Group; Health status index; Quality of life measures

Background

During the past decade there has been a burgeoning interest in the measurement of health-related quality of life, some of which has been motivated by a desire to widen the range of outcome measures used in clinical trials or other evaluative studies [12], some for auditing the performance of groups or individuals [10], and some for monitoring health levels in whole communities [4].

For these purposes a wide variety of methods has been used, depending on the interests of the measurers, their disciplinary background, the resources available

*The term 'EuroQol' is the copyright of the EuroQol Group.
**The 23 members of the EuroQol Group are listed in the Appendix.

to them, the context within which the measurements were to be made and used, and the level of accuracy required. For some, measurement simply meant accurate description, whilst for others it embraced also explicit valuation.

This diversity is both understandable and justifiable, since it will never be possible in such a broad field for one single all-embracing instrument to capture in an economical manner *everything* that *anybody* might want in *any conceivable* situation.

But so crowded is this field of endeavour becoming that there are a great number of rival approaches aiming to do very similar things, with little or no attempt at systematic comparison of their respective strengths and weaknesses either at a conceptual or at an empirical level. Moreover, such systematic comparison would be a formidable undertaking, for a mere half-dozen such instruments would generate 15 possible pairwise comparisons.

One obvious solution to this problem would be to compare each instrument only with the 'gold standard' amongst its class. But the trouble is that in this field there is no such gold standard, because initial decisions on structure and content made at the design stage were guided by the specific purpose of each instrument and compatibility with other instruments was not usually given any consideration.

The EuroQol Group

It was to create such a compatible common core of quality-of-life items that a small group of researchers from centres in five European countries, who shared a common interest in the valuation of health-related quality of life, met for the first time in May 1987 to draw together their respective expertise and experience.

Aims

The principal aim of what became known as the 'EuroQol Group' is to test the feasibility of jointly developing a standardised non-disease-specific instrument for describing and valuing health-related quality of life. It was intended that the instrument should complement other forms of quality of life measure, and should facilitate the collection of a common data set for reference purposes. Of particular importance to the group was the capacity to generate cross-national comparisons of health state valuations. It was envisaged that this in turn would facilitate the exchange of data on methods, and lead to standardisation in the collection and reporting of quality-of-life data.

Membership of the group

The EuroQol Group has a multidisciplinary membership (drawn from the fields of economics, mathematics, medicine, nursing, philosophy, psychology and sociology), and is made up of researchers who are actively engaged in practical experimentation. They share a common research interest in the evaluation of health

care services and meet on a regular basis. The present membership of the group is listed in the Appendix to this paper. The existence of the group has become widely known through personal networks and we feel that it is now time to disseminate some of the early fruits of our collective endeavour more generally. Hence, this paper aims to record both the objectives of the EuroQol Group and to report on its preliminary findings after some 3 years of successful collaboration.

Development strategy

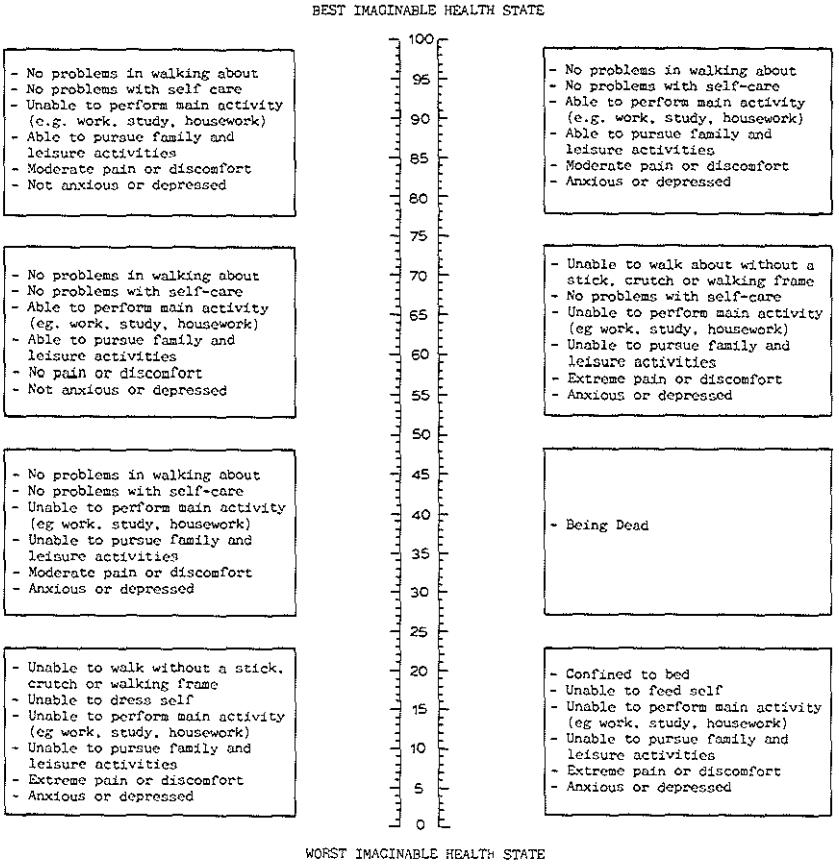
It was recognised from the start that a common core data set would not of itself be a comprehensive measure of health-related quality of life, but it would need to cover (in a selective manner) most of the different domains or dimensions which are frequently featured in such work. There was evident tension from the beginning between the desire to cover all the things that other people included, and the recognition that the more comprehensive the common core became, the less likely it was that collaborating research groups would have the resources available to collect the data in addition to whatever else they were doing.

We have also been motivated by another important strategic consideration. Within the field of health-related quality of life measurement there are two (often opposed) schools of thought. Both agree that quality of life is a multidimensional phenomenon, but they disagree about the implications of this. For one school of thought, this multidimensionality is held to imply that the measurement of quality of life must necessarily be multidimensional, so that the best we can hope for is the creation of a quality of life *profile*, in which measurement is only possible *within* a dimension or domain. Adherents to this way of thinking hold that measurement between dimensions or domains is not possible [3,11]. The second school of thought notes that since people have to weigh up the very diverse attributes of health to determine which, on balance, seems best, it should be possible to elicit such overall valuations by some suitable investigatory method which generates a single index value for each health state [2,9].

The EuroQol Group is grounded firmly in the second school. A multidimensional profile can be very useful in specific circumstances, for example in clinical trials where detailed information may be required. However, there are situations in which a profile is less useful and a single index is required. This is especially so in evaluative studies, for example in assessing cost-effectiveness. The same is true *a fortiori* when outcome measures are needed to help establish priorities across a wide range of health-care activities which bear on quality-of-life in very different ways.

The final general consideration which influenced our development strategy was ease of data collection. Health-related quality of life data is sometimes collected by highly skilled research teams, who are able to spend a great deal of time with each subject. On other occasions data is acquired by less sophisticated means, for example by postal questionnaires where interviewer follow-up is not feasible. Thus any supplementary data required for comparative purposes as part of our 'common core' must be capable of being collected under those (least favourable) circumstances.

Extract from the EuroQol questionnaire



What then were the properties which this new instrument needed to possess? It was decided early on that the instrument should be capable of being used in large-scale surveys of the community and that this necessarily meant that the instrument had to be designed as a self-completed questionnaire, probably for use in postal surveys. It needed to be sufficiently short so that it did not present a significant experimental burden on subjects and thereby jeopardise completion and/or return (especially when it is being used as an additional item in an already demanding research design). Since the ability of subjects to make consistent valuations of health states was not known, some form of test-retest was included in the design of the questionnaire. The single most significant property which was designed into the instrument by the EuroQol Group was a capacity to yield a single index value for any given health state.

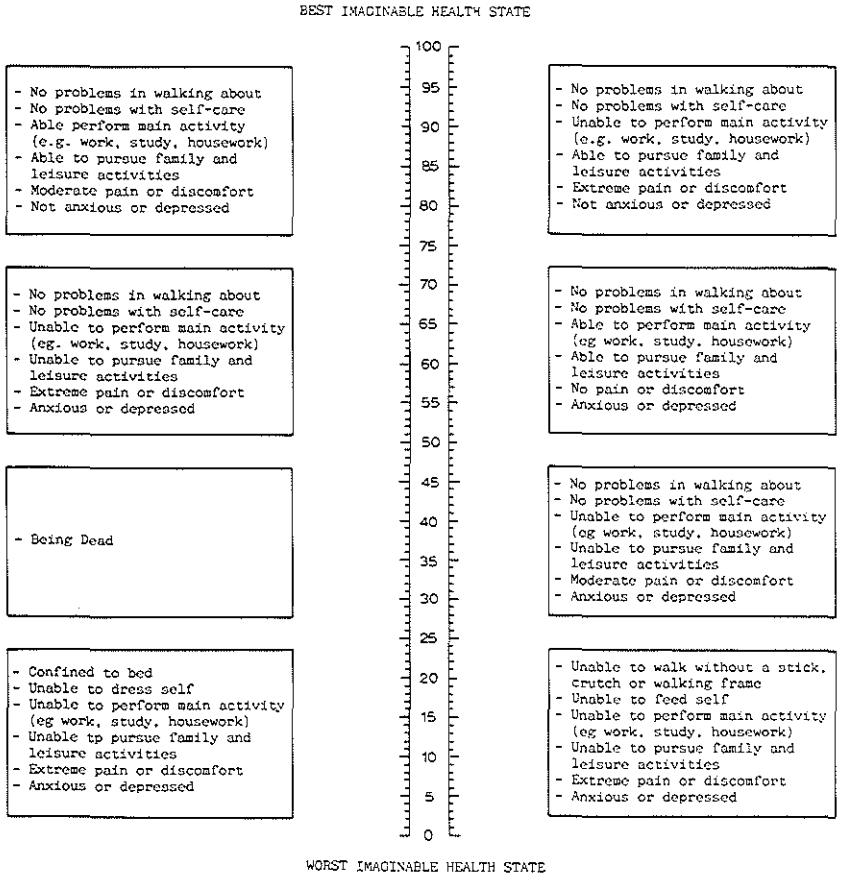


Fig. 1 Please indicate how good or bad these health states are, by drawing a line from each box to a point on the scale.

The EuroQol instrument

During the development phase we have been using a descriptive system which embraces six distinct dimensions or domains, with two or three categories within each. These dimensions were selected following a detailed examination of the descriptive content of existing health status measures including the Quality of Well Being Scale [6], the Sickness Impact Profile [1], Nottingham Health Profile [5] and the Rosser Index [7]. Additional measures currently being used by members of the EuroQol Group were also included in this meta-analysis. The resultant descriptive system is set out in Table 1 and defines a theoretical universe of 216 health states.

Table 1
EuroQol Descriptive Classification

Mobility
1. No problems walking about
2. Unable to walk about without a stick, crutch or walking frame
3. Confined to bed

Self-care
1. No problems with self-care
2. Unable to dress self
3. Unable to feed self

Main activity
1. Able to perform main activity (e.g., work, study, housework)
2. Unable to perform main activity

Social relationships
1. Able to pursue family and leisure activities
2. Unable to pursue family and leisure activities

Pain
1. No pain or discomfort
2. Moderate pain or discomfort
3. Extreme pain or discomfort

Mood
1. Not anxious or depressed
2. Anxious or depressed

In choosing this particular set of descriptors we were anxious to cover as many as possible of the domains frequently covered by others. We also wanted to cover a wide range of degrees of severity within each domain, so that the common core would offer scope for application to many different populations, from fairly healthy people living in their own homes and going about their usual activities, to severely ill patients in hospital.

In the preamble to the valuation task, subjects were given the following guidance: "We are trying to find out what people think about health. We are going to describe a few health states that people can be in. We want you to indicate how good or bad each of these states would be for a person like you. There are no right or wrong answers. Here we are interested only in your personal views."
Each subject was asked to rate 16 states, which were presented in two groups of eight across successive pages of the questionnaire (as shown in Fig. 1).

The choice of 16 states from amongst the 216 logically possible ones was dictated by the following considerations:
(a) We wanted some check on internal consistency, so we repeated two states from the first page on the second page. It will be noted that one of these was the state of 'being dead', so this left us with 14 living states to select.
(b) We wanted to include states which we knew (from earlier work) were likely to occur fairly frequently in practice.
(c) We wanted to include states representing a wide range of degrees of severity (which clashed to some extent with (b), since the more severe states are not amongst those that occur most frequently).

Methods and Results

The detailed work carried out during the developmental phase will be reported by individual researchers within the EuroQol Group in due course. It concerned variations in the instrument itself, alternative ways of eliciting valuations for the state dead, and comparisons with other valuation methods. However, one set of results emerging from this phase is of such intrinsic interest that it is worth disseminating immediately.

As noted earlier, one of our objectives has been to generate data on a comparable basis in different countries, so that we can check whether people have similar health experiences and whether they hold similar values about what it is like to be in various health states. We have devoted a great deal of attention to this latter issue. All collaborating centres have piloted the EuroQol questionnaire and collected data on health state valuations using a visual analogue scale based on the thermometer shown in Fig. 1. This form of scaling has been used elsewhere to derive values for health states [8].

Three somewhat more extensive studies were conducted in the U.K. (York), in The Netherlands (Bergen op Zoom) and in Sweden (Lund). In August 1988, 1321 questionnaires were mailed to a random sample of the adult patients of a GP group practice in Frome (U.K.). The return rate was 39% (n=522), of which 59% (n=310) provided fully usable responses. The Bergen op Zoom survey was conducted in December 1988 among a random sample of 200 heads of household. Sampling was based on postal codes. The return rate was 57% (n=112), of whom 66% (n=74) had fully completed the questionnaire. The third study was conducted in Lund among 1000 individuals selected at random from the general Swedish population. The

Table 2
Results of three pilot studies conducted with the EuroQol instrument in Lund (Sweden, n=208/1000), Frome (U.K., n=310/1321) and Bergen op Zoom (The Netherlands, n=74/200)
Note: The health state codes are described in Table 1, and the dimensions (mobility; self-care; etc.) to which each digit refers are in the same order as stated there.

Health state	Median valuations			Mean valuations			Standard deviation		
	Lund	Frome	BoZ	Lund	Frome	BoZ	Lund	Frome	BoZ
111111	100	99	95	93	95	93	13	10	13
111121	86	84	86	83	81	81	16	14	19
111112	75	70	75	69	67	71	21	18	22
111122	70	68	70	64	65	69	20	17	21
112121	65	70	65	61	67	63	22	18	23
112131	50	59	60	51	56	56	21	19	22
112222 (a)	35	40	43	36	41	43	20	17	21
112222 (b)	39	40	40	38	40	41	19	16	21
112232	35	35	33	36	36	37	20	17	23
212232	22	25	20	26	26	26	20	16	20
222232	10	10	7	14	12	12	19	12	15
232232	7	5	6	12	8	10	19	9	16
322232	4	2	5	9	5	10	18	7	18
332232	1	1	4	8	4	7	19	6	12
being dead (a)	0	0	3	10	10	19	24	20	25
being dead (b)	0	1	2	10	10	18	23	21	25

return rate amounted to 35% (n=349), of which 60% (n=208) were fully usable. In general, there seems little danger of selection bias as valuations vary little with background variables or response times.

The results of the three studies are summarised in Table 2. It can at once be seen that values for the majority of states are closely related, both in terms of their absolute values and in terms of their relative position within the samples. If the values are treated as ordinal data then Spearman's rank correlation coefficient can be used to quantify the pairwise relationship between the ordering of states in each study. In each case Spearman's rho was close to 1.0. Kendall's coefficient of concordance, W, for the rank ordering of states across all three studies was high and significant (W= 0.984, $p < 0.001$), indicating broad agreement regarding the ranking of states.

If the data are treated as cardinal values, then a regression analysis can be conducted. Where dependent and independent variables correspond perfectly the gradient of the regression line would be 1, with a zero residual constant. Regression analysis of the mean values in the BoZ and Frome studies yields an X coefficient of 1.06, and a constant of −4.80. The corresponding values for Lund/Frome are 1.04 and −2.04, and for BoZ/Lund are 1.01 and −2.37. The value for R^2 in all three cases is very close to 1.

These procedures were repeated using the median valuation as measure of central tendency and yielded the same results. We therefore conclude that whether treated as ordinal or cardinal data, the results of the three studies were strikingly similar.

Conclusions

We have been greatly encouraged by the progress we have made so far in achieving our principal aim of testing the feasibility of developing a new instrument for describing and valuing health-related quality of life, to serve as a linkage tool in a variety of different circumstances. We are also encouraged by the immediate spin-off benefits, such as the empirical data reported above. There is still some distance to go, however, and our system is not yet in its final form, since we are still experimenting with the form of the thermometer, with the descriptors, and with the treatment of 'being dead'. We have also not yet finally resolved the precise interpretation of our numbers in relation to other indices of health, derived by other means for other purposes. And although we already have our existing instrumentation available in Dutch, English, Finnish, Norwegian and Swedish, we are conscious of the fact that these countries do not constitute the whole of Europe and that our existing findings may not stand up if tested on populations with somewhat different cultural backgrounds (such as the Mediterranean countries).

Each of us has financed our own research input, and we have shared our empirical data freely with each other, and we propose to continue on this same collaborative basis. We would be pleased to hear from interested researchers in this field who would be willing to help us to extend this work in a practical way, and especially by offering 'test-bed' facilities whereby data, such as that reported here, could be

generated from other populations. Such contacts may be channelled through any of the members of the EuroQol Group, but preferably through one of the liaison officers, who are marked thus * in the list of members in the Appendix.

Appendix

Membership of the EuroQol Group

Health Economics Research Group
Brunel University
Uxbridge, Middlesex
England
Martin Buxton*
Moira O'Hanlon
Julia Rushby

Academy of Finland
Markku Pekurinen*
National Public Health Institute
Elimäenkatu 25 A
SF-00510 Helsinki

Harri Sintonen
University of Helsinki
Mannerheimintie 172
SF-00300 Helsinki

Swedish Institute for Health Economics
PO Box 1207
S-22105 Lund, Sweden
Stefan Bjork*
Bjorn Lindgren
Ulf Persson
Richard Brooks

National Institute of Public Health,
Geitmyrsveien 75,
0462 Oslo 4, Norway
Erik Nord*

Department of Academic Psychiatry,
Middlesex Hospital
Mortimer Street,
London, W1N 8AA, England
Richard Allison
Carole Butler
Rosalind Rabin*
Rachel Rosser
Caroline Selai

Erasmus University Rotterdam
P.O.Box 1738
3000 DR Rotterdam,
The Netherlands
Gouke Bonsel*
Marie-Louise Essink-Bot
Ben van Hout
Institute for Medical Technology Assessment
Department of Public Health and Social Medicine

Frank de Charro
Department of Economics
Faculty of Law

Centre for Health Economics,
University of York,
York, Y01 5DD, England
Claire Gudex
Paul Kind*
Alan Williams

*EuroQol liaison officers for each research group.

References

1 Bergner, M., Bobitt, R.A. et al., The Sickness Impact Profile: Conceptual Formulation and Methodology for the Development of a Health Status Measure. International Journal of Health Services, 6 (1976) 393–415.
2 Fanshel, S. and Bush, J.W., A Health Status Index and its Application to Health Services Outcomes. Operations Research, 18(6) (1970) 1021–1066.
3 Hunt, S.M. and McEwen, J., The Development of a Subjective Health Indicator. Sociology of Health and Illness, 2 (1980) 231.
4 Hunt, S.M., McEwen, J. and McKenna, S.P., Measuring Health Status: A new tool for clinicians and epidemiologists, Journal of Royal College of General Practitioners, 35 (1985) 185–188.
5 Hunt, S.M., McEwen, J. and McKenna, S.P. (1986) Measuring Health Status, London, Croom Helm.
6 Patrick, D.L., Bush, J.W. and Chen, M.M., Toward an Operational Definition of Health. Journal of Health and Social Behaviour, 14 (1973) 6–23.
7 Rosser, R.M. and Watts, V.C., The Measurement of Hospital Output. International Journal of Epidemiology 1 (1972) 361–368.
8 Sintonen, H., An Approach to Measuring and Valuing Health States. Social Science and Medicine, 15 (1981) 55–65.
9 Torrance, G.W., Measurement of Health State Utilities for Economic Appraisal. Journal of Health Economics, 5 (1986) 1–30.
10 Torus, C.D., Kaplan, R.M. and Atkins, C.J., The Costs and Effects of Behavioural Programs in Chronic Obstructive Pulmonary Disease. Medical Care 22(12) (1984) 1088–1100.
11 Ware, J.E., Brook, R.H. et al., Overview of Adult Health Status Measures Fielded in Rand's Health Insurance Study. Medical Care, 17(7) (1979) Supplement.
12 Wenger, N.K., Mattson, M.E. et al., Assessment of Quality of Life in Clinical Trials of Cardiovascular Therapies, American Journal of Cardiology, 54 (1984) 908–913.

Valuation of health states by the general public: feasibility of a standardized measurement procedure

Marie-Louise Essink-Bot
Gouke J. Bonsel
Paul J. van der Maas

Social Science and Medicine 31 (1990) 1201-1206
Published with permission of the editor

VALUATION OF HEALTH STATES BY THE GENERAL PUBLIC: FEASIBILITY OF A STANDARDIZED MEASUREMENT PROCEDURE

MARIE-LOUISE ESSINK-BOT,[1] GOUKE J. BONSEL[1,2] and PAUL J. VAN DER MAAS[1]

[1]Department of Public Health and Social Medicine and [2]Institute for Medical Technology Assessment, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Abstract—In the context of an international collaborative study we tested the feasibility of a utility measurement procedure in the Dutch general population. A postal questionnaire was sent to a random sample of 200 households in a town in the Netherlands (±50,000 inh.). Respondents were requested to value 14 six-dimensional health states by means of visual analogue scaling (VAS). The response was satisfactory (57%), though about a fifth of those willing to complete the questionnaire did not manage to use a VAS to express their opinion. Inconsistent answers occurred relatively seldom. Generally consensus existed with regard to relative (ranking) and absolute values of different health states.

These first results encourage us to go on with the development of this international instrument for the valuation of health states. We conclude the present instrument to be a feasible tool for valuation research, although questions relating to its measuring properties, as well as its reliability and validity remain to be answered.

Key words—utility measurement, valuation of health states, quality of life

INTRODUCTION

Cost-effectiveness analysis and cost-utility analysis are currently recognized as sources of information for decisions about the incorporation of new medical technologies in health insurance schemes [1, 2], e.g. governmental decisions about the reimbursement of heart transplantation in the United States, the United Kingdom and the Netherlands were made utilizing the results of national studies on the costs and effects of this intervention [3–5].

Despite a still growing interest in the measurement of costs and effects in health care and the appearance of standard textbooks, currently published studies still display large disparities regarding conceptual background and operational design, especially for measurement of effectivity [1, 6]. In cost-utility analysis, the measure of effectiveness consists of a factor representing length of survival and an index regarding health related quality of life. A way to find values for this index is by means of measuring health state preferences [6].

With regard to the measurement of utilities, at least two aspects contribute to observed variations in theory and practice. First, cost-utility analysis is an analytical tool derived from economic science. This leads to the implicit assumption of health being a commodity like other goods, to be acquired on a market [1, 7]. In fact, health does not fit perfectly into the definition of a commodity. Until now, health as a commodity is only operationally defined, using, e.g. the concept of quality adjusted life years (QALY's) or the more refined healthy-years equivalents [8].

In addition, the operational designs of various empirical techniques to determine the quality adjustment index show many differences.

One might argue that these disparities are a characteristic of an immature branch of science; the lack of a paradigm would be only temporary [9]. Even if this is true, there are several undesirable consequences of this situation.

The lack of standards hampers scientific judgement, especially of aspects of validity [1]. Incomparability of results also prevents governments, health insurance companies and others from setting priorities among different technologies, based on information from cost-effectiveness analysis or cost-utility analysis. Finally, the application of different methods limits the possibility to benefit from studies already conducted in other countries.

Following an initiative of Professor Alan Williams (York) in 1987, several European research groups joined their research efforts aiming at two goals: first, development of a common core of methods and practical devices to measure health state preferences, and second, establishment of a common set of data collected with these instruments in different European countries. Details on the aims and the development of this research group are to be published elsewhere. Based on the shared experience and on intermediate results of pilot-studies, consensus was reached about a prototype of a common utility measurement instrument in 1988. An international pilot-study with this measuring instrument was conducted since.

This article describes the first results of the Dutch part of the international pilot-study, addressing the following questions:

1. Do the results of this pilot-study indicate possibilities to develop the present instrument into a feasible instrument for large scale surveys? This question should be focused on the following aspects:
   —the feasibility of the valuation of complex multi-dimensional health state descriptions. The method that is used here consists of six

dimensions. This is opposite to, e.g. the approach of Rosser and Kind who use health state descriptors composed of two dimensions, i.e. disability and distress [10]. The Rosser and Kind approach simplifies the task of valuating health states, but may oversimplify the concept of health;
—the feasibility of the valuation of health states by the general public, as opposed to the usually well-educated subjects in the literature [10, 11];
—the feasibility of the valuation of health states by means of a postal questionnaire, as opposed to the common interviewer-supported designs [12].

2. What is the actual value of health states, ranging from the health state of, e.g. heart transplantation candidates to that of the healthy population?
3. Do the results of this pilot-study indicate the existence of consensus about the rating of health states among respondents to a degree justifying future research to be directed at application of the results in cost-utility analysis?

Questions concerning the influence of background variables (including nationality) on the valuation of health states, validity of the measurement procedure and use of the results in cost-utility analysis will be addressed more extensively in later papers, combining the results of several national studies.

METHODS

Design of a utility measurement procedure

Utility measurement consists of four consecutive steps:

1. health status measurement of the target population, usually patients subject to some intervention;
2. translation of resulting health state data into verbal descriptions of the health states in the target population, sometimes called 'scenario's' [11, 13];
3. valuation of these 'scenario's' by a panel consisting of representatives of the general public, experts or patients [11, 13];
4. implementation of the results into a cost-utility analysis, preferably combining effects on survival with that on health status (a procedure yielding, e.g. QALY's) [11].

The concept of health used

Defining health status, a multi-dimensional approach was adopted in order to take the complexity of the concept of health into account. A concept of health status, defined by six dimensions (apart from survival state), to be used in step 1–3 was agreed on.

Table 1. Dimensions of the health concept

Mobility (3 levels)
Daily activities and self care (3 levels)
Work performance (2 levels)
Family and leisure performance (2 levels)
Pain/discomfort (3 levels)
Present mood (2 levels)

Table 2. Example of a complete scenario

State 112232:
No problems in walking about
No problems with self-care
Unable to perform main activity
Unable to pursue family/leisure activities
Extreme pain or discomfort
Anxious or depressed

The choice of dimensions was suggested by existing health questionnaires such as, e.g. the Nottingham Health Profile and the Sickness Impact Profile, and by the health concepts used by Patrick and Bush and by Rosser and Kind respectively [11, 12, 14, 15]. Within each dimension the state of a patient should be described by one out of a set of dimension-specific descriptions. Three dimensions were subdivided into three categories, the other three into two categories (see Table 1). Thus a complete health state descriptor or 'scenario' consists of six dimension-specific statements ('items'). Theoretically this set of dimensions and items allows for 216 ($2^3 \times 3^3$) permutations. Each possible scenario can be characterized by a string of the values of the item-levels on each dimension, '1' representing the optimal level, '2' respectively '3' representing the worst level. One example of a scenario is shown in Table 2.

The questionnaire

A questionnaire was developed for the valuation of scenario's (step 3 of the aforementioned procedure) by members of the general public.

For our study 13 scenarios were selected both on empirical grounds (counting health state frequencies in patient data from other studies) and on theoretical grounds [16, 17]. The state of 'being dead' was added [18]. Two states were presented twice, resulting in 16 health states in the questionnaire. The scenarios were presented on two pages of the questionnaire. A rectangular box was drawn around each scenario and four boxes were placed on either side of a vertically placed visual analogue scale (VAS) in a random sequence, as illustrated by Fig. 1. The endpoints of the scale were marked with the words 'worst imaginable health state' (0) and 'best imaginable health state' (100). No additional information about the interpretation of the scale was added. The duration of each state was supposed to be 1 year; what happens afterwards was stated not to be known. To avoid misinterpretation a detailed instruction paragraph was added, including illustration of the method of valuation by a non-health-problem-related example. Also the lay-out of the questionnaire was carefully designed.

The main task of valuating the health states was preceded by the task of classifying and valuating one's own health state at the time being. Questions about characteristics that might influence the rating of health states (a.o. age, educational level, experience with illness, own health state) were presented at the end of the questionnaire.

The sample

The questionnaire was mailed to a random sample ($n = 200$) of the general population in December

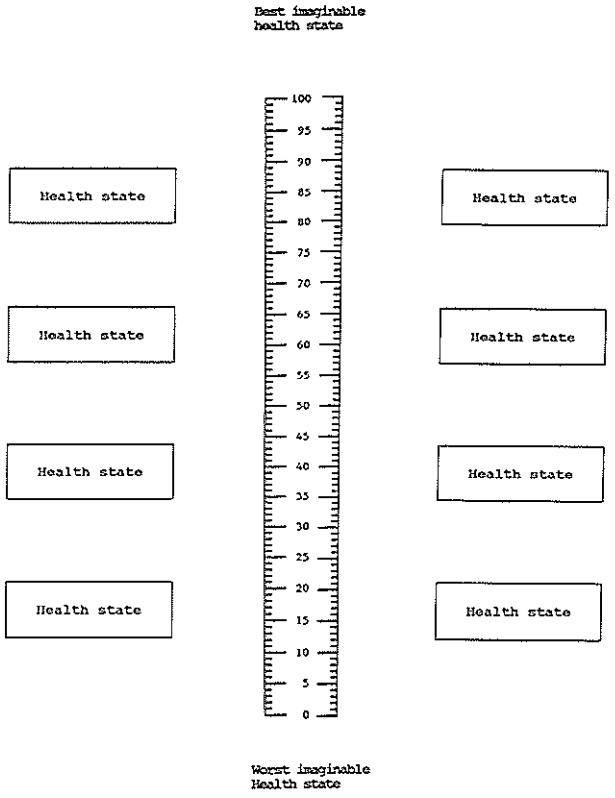


Fig. 1.

1988, followed by a reminder 2 weeks later. Sampling was based on postal codes.

RESULTS

Response was obtained from 112 persons. The response rate (excluding 4 dead persons) was 57%, assuming all addresses were provided correctly, which is obviously optimistic. Background data of the respondents are shown in Table 3. Comparison with data from the sampled population shows some differences in age and sex distribution. The relative overrepresentation of men among the respondents may be due to the method of sampling, as the questionnaire was directed to the administrative head of each household.

Five respondents returned their questionnaire blank. Twenty-one (20%) of the remaining 107 clearly had not understood the task, i.e. the use of VAS to express their opinion. This response was chiefly found among the older and less educated respondents [chi-square: age (2 strata) $P = 0.003$, education $P = 0.001$]. These respondents were left out of further analysis. Five of the remaining 86 completed only one of the two pages on which the valuation task was presented, overtly because they thought the second one to be a double from the first. Data from these five were included.

To assess aspects of the feasibility of the questionnaire, respondents were questioned about the difficulties they experienced answering it. Forty-three percent of the respondents judged the questionnaire as being very (6%) or rather (37%) difficult, while 57% reported it to be fairly (45%) or very (12%) easy. The mean time required to complete the questionnaire was 20.3 min (SD 12.4 min, $n = 86$).

Respondents rated their own health status on the VAS (range: 0–100) with a mode of 85, a median of 85 [interquartile range (IQ) 8] and a mean of 81 (SD 18). Indeed, those who classified themselves on all predefined dimensions as being in the best category (111111) ($n = 52$), attributed a significantly ($P < 0.01$) higher value to their own state of health (mean 89, SD 7) than those who reported a suboptimal level on any dimension ($n = 40$; mean 70, SD 22).

Table 3. Background data of respondents ($n = 86$), excluding inappropriate* response; some background data of the sampled population (age ≥ 15)

Variable	Appropriate response	Sampled population (%)
Age:		
15–29	16 (19%)	30
30–44	32 (37%)	28
45–59	11 (13%)	19
60–74	22 (26%)	15
≥75	5 (6%)	7
Sex:		
Male	54 (63%)	52
Education:		
Minimum schooling	37 (43%)	55
Intermediate	34 (40%)	31
Higher/degree level	15 (17%)	14
Missing	0 (0%)	
Main Activity:		
Employed	46 (54%)	
Retired	19 (22%)	
Housework	15 (17%)	
Student	1 (1%)	
Incapacitated	4 (5%)	
Seeking work	1 (1%)	
Missing	0 (0%)	
Rating of own health:		
<80	18 (21%)	
80–90	29 (34%)	
>90	32 (37%)	
Missing	7 (8%)	

*For definition see text.

The results of the valuation of the 16 elected health states are summarized in Table 4. The ordering of the states is according to the medians, as at least ordinal measuring level may be assumed. Ordering according to the arithmetical means does not result in any change in ranking, however.

The dispersion of the attributed values is rather large, especially for 'bad' health states. 'Being dead' yields a heterogeneous response, the range of attributed values ranging from 0 to 100. Seven of 80 respondents (9%) valuated 'being dead' equal to or higher than 50. Two health states were presented twice (112222 and 'being dead'). Scores were compared for both pairs. Individual correlation was high

Table 4. Valuations for 14 health states

Health state*	Mode	Med.	I.Q.†	Mean	SD	n
111111	100	95	5	92	14	86
111121	80	85	10	81	19	82
111112	85	78	10	73	21	81
111122	70	70	13	69	21	86
112121	60	65	15	64	22	85
112131	65	60	14	55	23	83
112222 (a)‡	30	43	13	42	21	86
112222 (b)	40	40	13	41	21	82
1122328	25	33	11	37	22	81
212232	20	20	8	26	20	86
Being dead (a)‡	0	5	23	21	26	80
Being dead (b)	0	5	20	19	25	77
222232	5	8	6	12	15	85
232232	0	6	4	11	16	83
322232	0	5	5	10	17	81
332232	0	4	5	7	12	85

*For clarification of representation of health states by strings of numbers see Methods and Tables 1 and 2.
†Med. = median score, I.Q. = interquartile range.
‡These two states ('112222' and 'being dead') were presented twice in the questionnaire.
§The string '112232' represents the scenario presented in Table 2.

Table 5. Consistency according to item-rank differences between pairs of health states

Distance	Pairs	Inconsistent (%)
1	14	16.3
2	14	5.9
3	13	2.8
4	12	1.8
5	9	1.7
6	8	1.7
7	7	1.4
8	4	1.2
9	1	1.2
Total	82	5.0

(Spearman's rank correlation coefficient 0.76 and 0.95 respectively; Pearsons's correlation coefficient 0.69 and 0.94 respectively), though small, insignificant ($P > 0.05$ for both pairs) differences of the means were found (see Table 4). The influence of background variables (e.g. age, sex, level of education) on the resulting valuations appeared to be insignificant.

Next we tested the logical consistency of valuations, both on a group level and on an individual level. Consistency of valuations may be checked by comparison of values attached to pairs of scenario's which can be logically ordered according to their description. For instance 112222 might be expected to show a higher score than 212232. When a dimension description varies in two directions, e.g. 112131 and 112222, with in this case the first scenario on two dimensions better and one dimension worse, no logical rank order can be established. 'Being dead' was excluded from this analysis. The median answers (group level) do not exhibit any inconsistency (see Table 4). However, on an individual level inconsistencies occurred. The 14 health states account for 82 pairs of logically ranked scenario's. Overall, only 5% of the answers proved to be inconsistent on the individual level. Illogical ranking occurred more often as pairs of health states were more alike (see Table 5).

The next item to be addressed is consensus between respondents. This was first tested by comparing individual rankings with the group ranking, as the measuring level is assumed to be at least ordinal. As cardinal measuring properties are conceivable, a comparison of individual values with the group means was also carried out. The results are shown in Fig. 2.

DISCUSSION

The feasibility of a measurement instrument to elicit valuations for various complex health states from the general public was investigated. The device of this instrument was based on previous experience and supporting pilot studies by a collaborative group of European researchers [19–21]. Essentially it consists of valuation by VAS of six-dimensional health state descriptions in an unsupported situation (postal questionnaire). Some of the arguments favouring its feasibility are:

—the relatively high response, taking into account the demanding character of the questionnaire;
—the acceptable level of experienced difficulty;
—the interpretability of the results.

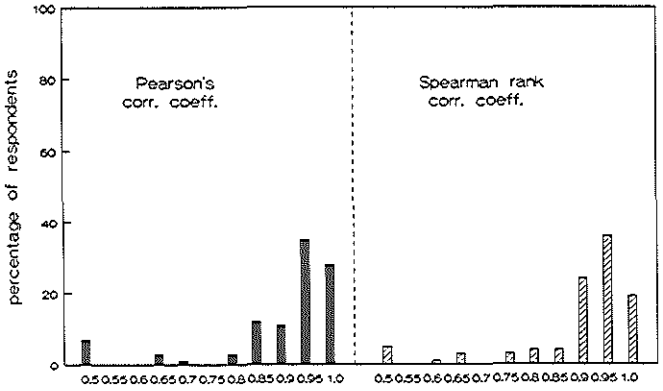


Fig. 2. Correlation of individual scales with group scale ($n = 74$).

Thus rating of health states on a VAS by postal questionnaire appears to be feasible. However, inappropriate response of those returning the questionnaire did occur (20%), and was related to age and level of education. This is probably of little importance, as further analysis of our data did not reveal significant influence of the level of any of the sociodemographic variables. Improvement of the instruction paragraph probably will enhance completion rates. A thorough non-response study will be undertaken to answer the question directly whether persons who would generally not respond differ from respondents in their valuation of health states.

Judging from the response-quality and the logical consistency, the valuation task was generally well understood. Also the fact that respondents who *classified* their own health logically higher also *rated* it significantly higher, may lead to the conclusion that the laymen interpreted the judgement task correctly. The dispersion of the attributed values is fairly large. However, psychophysical methods typically yield highly variable observations [12].

The consensus found among respondents with respect to the ranking of the scenario's is of importance [22]. It is a support for the validity of the questionnaire, as well as for earlier evidence of homogeneity of society with regard to health state preferences [10, 22]. However, the possibility of biases introduced by the response rate should be kept in mind. A comparison of results of this survey with preliminary results in other European countries shows virtually the same ranking order of health states (data to be published).

Though these provisional results seem to warrant extended use of the measurement instrument, there remain additional questions to be answered. In the first place the reliability and the validity of the questionnaire should be determined. For the validity this concerns the content validity (are we measuring health state valuations?) as well as external validity (can we generalize results to subgroups of the population, e.g. patients, health care workers etc?).

Secondly the measurement level of the produced values should be established. Are respondents merely ranking the states, or are the numbers on the scale interpreted quantitatively? Research on measuring properties of a VAS in this context implies comparison with other valuation methods [time trade off (TTO), standard gamble (SG)], from which also conclusions about the validity of the measurement procedure (are we measuring health state preferences?) might be drawn. Methods like TTO and SG can not be operationalized on a postal questionnaire base, so a more structured interview format will be necessary. Also, results of the VAS with and without a structured interview setting should be compared.

For application of the health state valuations in cost-utility analysis, at least interval measuring level is necessary [6, 23]. If the health states appear to be only ranked by the respondents, a scaling procedure could be performed in order to achieve cardinal values. Even if the measurement level is interval, the values resulting from the measurement procedure should probably not be used directly as a quality index in the computation of QALY's. The relation between the elicited valuations to health state preferences as measured by time trade off or standard gamble may be, e.g. logarithmic or linear [6, 24, 25]. Probably a transformation procedure should be employed before use of the results in cost-utility analysis [26]. More generally, the way to apply the results in QALY-calculations should be determined.

Problems remain with the valuation of 'being dead'. Controversy exists whether it should be incorporated in a system of health states. For application in computations of quality-adjusted life-years (QALY's) it is not necessary to elicit values for 'being dead', as 'being dead' in this context means '0 life years'. Also presentation of 'being dead' as a health state encounters opposition of respondents. This might be the result of assignment of two conflicting meanings to 'being dead', viz. 'lack of life years' and 'something very bad'. In the last intepretation, 'being dead' serves as a point of reference on the scale. This double-interpretation might be the cause of cognitive problems in the comparison of 'being dead' with health states.

Another remaining question regards response patterns. We observed, e.g. 'dichotomous' response patterns: all states were valuated '100' except the very worst, or all states were rated '0' except some (nearly) optimal ones. Such response patterns may in fact reflect a cognitive difficulty to distinguish more 'shades' than 'healthy' and 'unhealthy'. Another possibility is that they reflect extremes of unknown individual distribution rules. We think this is the same phenomenon as Llewellyn-Thomas found in her analysis of standard gamble validity [13]. McNeil also describes this response pattern, as well as a formal way to apply such results as utility weights [27].

Routine implementation of an improved version of the present measurement instrument also requires systematical research on methods to describe health states of patients (steps 1 and 2). Utility instruments described earlier do not provide unequivocal algorithms to categorize patients [10, 11].

Though many questions are yet to be answered, the first results with this instrument are encouraging enough to continue the search for a universally applicable utility measurement instrument for health states.

REFERENCES

1. Drummond M. F., Stoddart G. L. and Torrance G. W. *Methods for the Economic Evaluation of Health Care Programmes*. Oxford University Press, 1987.
2. Leaf A. Cost effectiveness as a criterion for Medicare coverage. *New Engl. J. Med.* **321**, 898, 1989.
3. Evans R. W., Manninen D. L., Overcast T. D., Garrison L. P., Yagi J., Merrikin K., Jonsen A. R. *et al.* The National Heart Transplantation Study: Final Report. Battelle Human Affairs Research Centers, Seattle, WA, 1984.
4. Buxton M., Acheson R., Caine N., Gibson S. and O'Brien B. Costs and benefits of the heart transplant programmes at Harefield and Papworth Hospitals: Final Report. Department of Economics, Brunel University and Department of Community Medicine, University of Cambridge, 1985.
5. Charro F. T., Bonsel G. J. and Hout B. A. van. Costs and effects of heart transplantation. Vol. 9: Final report (in Dutch). Department of Economics, Faculty of Law, Erasmus University Rotterdam, the Netherlands, 1988.
6. Torrance G. W. Measurement of health utilities for economic appraisal. A review. *J. Hlth Econ.* **5**, 1–30, 1986.
7. Lipscomb J. Value preferences for health: meaning, measurement and use in program evaluation. In *Values and Long-Term Care* (Edited by Kane R. L. and Kane R. A.). Toronto, 1982.
8. Mehrez A. and Gafni A. Quality-adjusted life years, utility theory, and healthy-years equivalents. *Med. Dec. Mak.* **9**, 142, 1989.
9. Kuhn T. S. *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago, IL, 1962.
10. Rosser R. and Kind P. A scale of valuations of states of illness: is there a social consensus? *Int. J. Epidem.* **7**, 347, 1978.
11. Patrick D. L., Bush J. W. and Chen M. M. Toward an operational definition of health. *J. Hlth soc. Behav.* **14**, 6, 1973.
12. Kind P. and Rosser R. The quantification of health. *Eur. J. Soc. Psychol.* **18**, 63, 1988.
13. Llewellyn-Thomas H. A., Sutherland H. J., Tibshirani R., Ciampi A., Till J. E. and Boyd N. F. The measurement of patient values in medicine. *Med. Dec. Mak.* **2**, 449 1982.
14. Hunt S. M., McEwan J. and McKenna S. *Measuring Health Status*. Croom Helm, London, 1986.
15. Bergner M., Bobbitt R. A., Kressel S., Pollard W. E., Gilson B. S. and Morris J. R. The Sickness Impact Profile: conceptual formulation and methodology for the development of a health status measure. *J. Hlth Serv.* **6**, 393, 197–6.
16. Kind P. and Carr-Hill R. The Nottingham Health Profile: a useful tool for epidemiologists? *Soc. Sci. Med.* **25**, 905, 1987.
17. Erdman R. A. M., Bonsel G. J. and Mast R. C. van der. Psychological aspects before and after heart transplantation: 5 years experience. Paper presented at the *World Conference on Psychiatry*, Athens, 1989.
18. Sintonen H. An approach to measuring and valuing health states. *Soc. Sci. Med.* **15C**, 55, 1981.
19. Williams A. Economics of coronary bypass grafting. *Br. med. J.* **291**, 326, 1985.
20. Buxton M., Ashby J. and O'Hanlon M. Valuation of health states using the time trade-off approach: report of a pilot study relating to health states one year after treatment for breast cancer. Health Economics Research Group, Brunel University, Discussion Paper No. 2, London, 1986.
21. Bonsel G. J., Bot M. L., Boterblom A. and Veer F. van't. Costs and effects of heart transplantation. Vol. 2C: Quality of life before and after heart transplantation—Results (in Dutch). Institute of Public Health and Social Medicine, Erasmus University Rotterdam, the Netherlands, 1988.
22. Hilden J. The nonexistence of interpersonal utility scales. A missing link in medical decision theory? *Med. Dec. Mak.* **5**, 215, 1985.
23. Weinstein M. C. Economic assessments of medical practices and technologies (tutorial). *Med. Dec. Mak.* **1**, 309, 1981.
24. Llewellyn-Thomas H. A., Sutherland J., Tibshirani R., Ciampi A., Till J. E. and Boyd N. F. Describing health states. Methodologic issues in obtaining values for health states. *Med. Care* **22**, 543, 1984.
25. Stevens S. S. *Psychophysics*. J Wiley, New York, 1975.
26. Torrance G. W. Multiattribute utility theory as a method of measuring social preferences for health states in long-term care. In *Values and Long-Term Care* (Edited by Kane R. L. and Kane R. A.). Lexington, MA, 1982.
27. McNeil B. J. and Pauker S. G. Decision analysis for public health: principles and illustrations. *A. Rev. publ. Hlth* **5**, 135, 1984.

The need for liver transplantation

Gouke J. Bonsel
Marie-Louise Bot
Ids J. Klompmaker
J. Dik F. Habbema
Maarten J.H. Slooff

Manuscript 1988

Abstract

The need for liver transplantation (liver transplantation) in the Netherlands was determined as part of a medical technology assessment. We aimed at estimation of the need for liver transplantation following medical standards. Due to the absence of longitudinal studies of end-stage and congenital liver diseases three indirect methods were used.
The *practice-based* method extrapolates current transplantation activities, accounting for incomplete referral, relaxation of contraindications, inclusion of acute hepatic failure and retransplantations. The *mortality-based* method starts from national age and sex-specific mortality data on the relevant liver diseases. Adjusting factors were estimated to express the quantitative association between death due to a specific liver disease and suitability for liver transplantation. This method enables to estimate the impact of inclusion of older patients.
The *incidence-based* method applies available epidemiological data on living patients with specific liver diseases.
The three methods yielded corresponding results suggesting an annual need of at least 25 liver transplantations. Without change of upper age limit (60 year) and inclusion criteria (no alcoholic cirrhosis) this number may rise to a maximum of about 70 transplantations in the near future. The actual number of transplantations and the uneven distribution of referred patients suggests the existence of unmet need.
International application of Dutch results, i.e. a normative need for liver transplantation of 2 to 5 per 10^6 inhabitants, is only justified in countries with comparable epidemiological structure and similar selection criteria.

1 Introduction

Liver transplantation, though never assessed for efficacy in a controlled trial, is performed in a continuously growing number of centres all over the world[1]. As the costs of the procedure are impressive, ranging from $180.000 to $ 460.000, health insurance companies have been reluctant to incorporate liver transplantation into their insurance schemes[2,3,4]. In the Netherlands orthotopic liver transplantations

have been performed since 1979 in the Academisch Ziekenhuis of Groningen (AZG). The Dutch government and the Dutch health insurance companies agreed in subsidizing this liver transplantation-programme, excluding the procedure from regular reimbursement. From 1985 to 1988 a comprehensive medical technology assessment of liver transplantation was carried out to assess the costs and effects of the procedure and to determine the future need for liver transplantation[5]. The results would be used for health policy decisions on the national level.

This article presents three empirically based procedures to estimate the need for liver transplantation to answer the following questions:

- does the present situation show any unmet need for liver transplantation in the Netherlands ?
- what is the future need for liver transplantation in the Netherlands ?
- is generalization of our estimations of need to other countries justified ?

2 Definitions, methods and data

2.1 Patient definition

Criteria for patient selection are described in the protocol of the liver transplantation-programme and elsewhere[1,6,7]. Patients should suffer from some irreversible non-alcoholic liver disease without complications or accompanying diseases, which might affect transplantation results.

Most contraindications are relative i.e. their potential influence is balanced against other favourable and unfavourable circumstances. Further improvement of the results may decrease the weight of contraindications, leading to an increasing proportion of suitable candidates among patients with irreversible liver disease.

2.2 Definition of need

Three types of need* may be distinguished[8,9,10]:

- normative or medical need
- felt need
- latent need

Normative need is defined as the need for health services according to standards of professional care, if existent. It is related to the expected benefits of a particular health service. Felt need is defined as the need experienced by the population. Media and other sources of professional and non-professional information may exert great influence on felt need. If need is not accommodated sufficiently by supply of medical services, latent need exists. Need may be latent as defined by medical standards, from the viewpoint of a patient, or both. In our case we were primary interested in normative need.

A maximally valid estimation of the frequencies of particular disease-stages and their change over time on the national level, would imply a nation-wide longitudinal epidemiological programme. However, given limited resources other, indirect methods must be applied[10,11,12].

Evans estimated the 'true' need for heart transplantation in the United States with population-based epidemiological data and with utilization-based data ('true' need was defined nearly identical to normative need)[10]. The use of utilization-based data assumes a firm relation between need and supply. This method overestimates 'true' need if unjustified supply is substantial but underestimates 'true need' if latent need from a medical point of view is present.

Except for acute hepatic failure, we tried to estimate normative need of liver transplantation irrespective of availability of resources (number of centres, available donors and the degree of reimbursement). Both population-based epidemiological data and utilization-based data were used, supplemented by various data from the AZG.

---

* We do not take economically defined need into account.

2.3 *Methods for empirical estimation of normative need*

Three methods were used to determine the normative need: a practice-based, a mortality-based and an incidence-based method.

The *first* method takes the 10-year transplantation experience of the AZG (1978-1987) as point of departure. We regarded the annual volume of liver transplantations in the AZG to be a valid proxy for the normative need in the Netherlands provided that:

- patient selection has been unchanged in the AZG-programme
- the AZG-programme has not been limited by external constraints (financial resources, donors, incomplete referral, etc.)
- the transplantation-programme has been in a steady state

None of these conditions were perfectly satisfied. As a considerable average lag-time was observed between candidacy and transplantation, we used the annual number of definitively accepted patients as the basic proxy of future transplantation volume, which was subsequently refined for known sources of distortion.

The *second* method started from national mortality statistics and should be regarded as a population-based method. We assumed that in the absence of a liver transplantation programme, transplantation candidates may be found among those who died due to specific chronic liver diseases. We regarded the use of mortality data to be justified if:

- the medical condition for which the procedure is indicated can be properly classified according to the current International Classification of Deaths (ICD 9th revision)
- the interval between supposed application of the medical procedure and death due to the specific disease is reasonably small
- data are derived from a period, in which no effect of the application of the medical procedure may be expected

Though in the case of liver transplantation these three conditions seemed sufficiently be satisfied, sensitivity-analysis was carried out to estimate the effects of age limits, contraindications and incomplete referral. Transplantations for quality of life improvement or replacement of a non-functioning graft were accounted for separately.

The *third* method applies population-based epidemiological data on living patients with end-stage liver disease. This method needs additional information on the stage or severity of the particular disease. E.g. data on the incidence of histologically proven primary biliary cirrhosis may only be useful if data on the distribution of age, clinical stage (Child-Pugh classification), comorbidity, and clinical course are also available.

If information on the epidemiology of non-alcoholic liver diseases in the Netherlands was lacking, we projected epidemiological data from other countries on the Dutch population, provided that epidemiological comparability could be expected. Within Western Europe only comparability with the United Kingdom seemed sufficient for this purpose[13,14,15]. However, most data available originated from a comprehensive study of Vierling, who used data from American sources[16]. Eventually we compared the resulting projections with Dutch hospital admission data of 1982, provided by the Dutch registry of hospital activities (Stichting Informatiecentrum voor de Gezondheidszorg, SIG). No information about the readmission rate was available. With some diseases we compared the results of the incidence-based method with national mortality statistics.

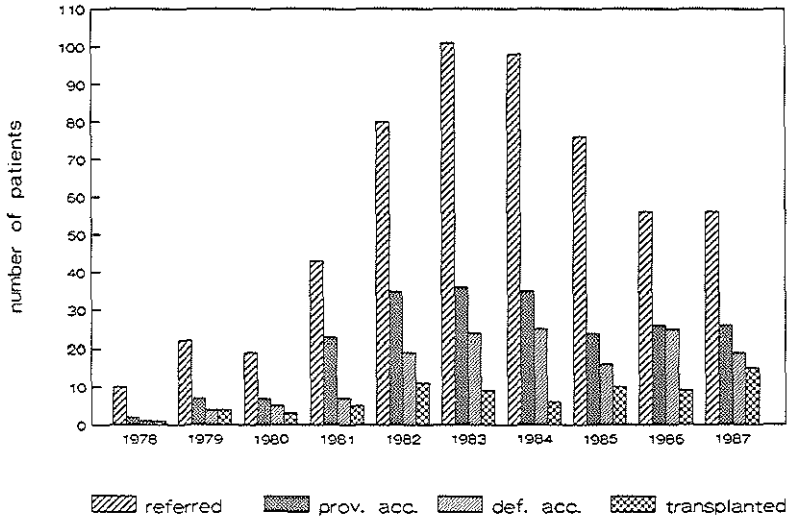
3 Results

All results are presented as normative need for liver transplantation for the Netherlands, and as normative need per 10^6 inhabitants. The minimum level of an adjustment factor reflects the current state of affairs, the maximum level shows the largest expected change.

3.1 Practice-based method

Figure 1 shows the annual inflow and throughflow of patients in the liver transplantation-programme. During 1978 to 1987 a total number of 145 patients was definitively accepted for liver transplantation. At the end of 1987 about half of these had been transplanted. Recently the annual number of definitively accepted patients has levelled off at about 20, which was taken as the basic number for further computations.

Figure 1 Annual number of patients in subsequent diagnostic stages. 1978-1987.



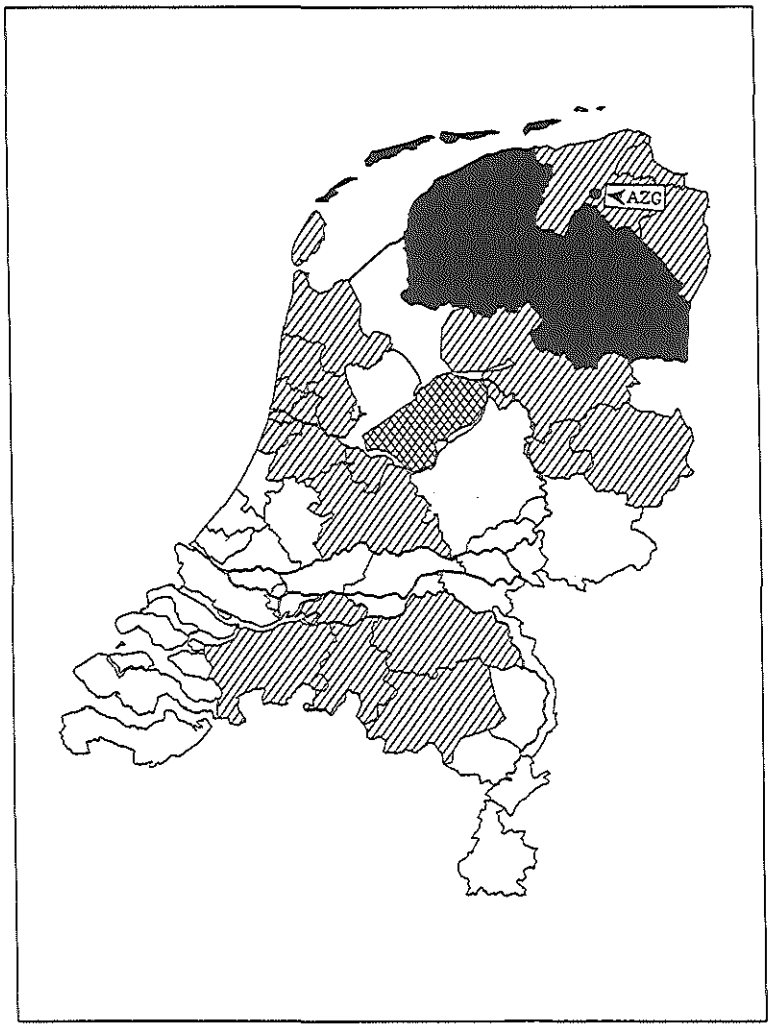
PROV.ACC. = provisionally accepted DEF.ACC. = definitively accepted

Below the correction factors are discussed, which were used to adjust this basic number for changing selection and suboptimal referral.

Relaxation of contraindications

Examination of the medical records of non accepted patients revealed 40 patients who might have been accepted when accepting policy would have been more liberal. In 34 cases liver dysfunction was considered too advanced to justify the risks which accompany the operation. An additional 6 HB_sAg-positive cirrhosis patients were rejected in accordance with the liver transplantation protocol. IIf all these patients are included a multiplication factor results of 1.3 (=185/145).

Figure 2 Regional distribution of definitively accepted patients for liver transplantation. 1978-1987



Referral patterns

Next we investigated referral patterns in the Netherlands by comparing the standardized number of definitively accepted patients in each province ('county'). The distribution of definitively accepted patients appeared to be rather uneven. Taking the four northern counties together, more definitively accepted patients originated from this part of the country as compaired to the remaining counties (see Figure 2 on previous page). If all provinces would show the same level of referral as the northern provinces, a total of 274 definitively accepted patients should be expected. Thus a factor with a maximum level of 1.9 (=274/145) should be applied to account for a complete change of the referral level of the western and southern provinces towards the level of the north. This estimation assumes no unjustified supply and no latent need in the north and no significant regional differences in the epidemiology of non-alcoholic liver diseases.

Additional diagnoses

We added an annual number of 2 to 10 patients with acute hepatic failure, based on the experience of all major centres in the Netherlands. The experience of the Rotterdam liver centre, which accepts this type of patients from all over the country for auxiliary partial liver transplantation, suggests that only in small number of cases of acute hepatic failure transplantation can be performed [17,18,19].

Retransplantations

The actual retransplantation rate is about 11%. Different centres have rapported retransplantation rates ranging from 10 to 25 %[20]. An upper limit of 15% was chosen, based on European experience[21].

Final result

Table 1 shows the final results of the practice-based method suggesting a normative need between 1.7 and 4.9 per 10^6 inhabitants. Chronic cirrhosis constitutes about two thirds of this number. Clearly effects of changes of the current age limit can not be shown with this method.

Table 1 Practice-based estimation of normative need for liver transplantation in the Netherlands.

	Minimum	*Maximum*
Definitively accepted patients (annual, AZG)	20	20
Adjustment Factors		
- Relaxation contraindications	x 1.0	x 1.3
- Even referral-patterns	x 1.0	x 1.9
	20	49
- Additional indications	+ 2	+ 10
	22	59
- Retransplantations	x 1.10	x 1.15
Total per year in the Netherlands	24	68
Normative need per year per 10^6 inhabitants	1.7	4.9

3.2 Mortality-based method

National mortality data were made available by the Central Bureau of Statistics (CBS). Registration of causes of death in the Netherlands offers complete coverage of the country and is based on the ICD 9th revision[22]. Table 2 shows the selected disease-groups, together with the corresponding I.C.D.-codes and the average annual number of deaths in the Netherlands in the period 1979-1984.

A problem exists with regard to ICD 571.5, 571.8, and 571.9. These codes are assigned to deaths due to cirrhosis or chronic liver disease without any indication of a specific cause. As the number of deaths in category 571.5 accounts for 20% of the total number of deaths in ICD 571, the contribution of non-alcoholic deaths to ICD 571.5, 571.8 and 571.9 should be estimated. After consultation of experts, application of the proportion of non-alcoholic among all specified deaths in ICD 571 seemed justified (19.6 / (239.7+19.6) = 0.08; rounded to 0.10). Nevertheless computations were repeated with a supposed non-alcoholic proportion of 0.25. Acute hepatic failure were excluded from analysis (see practice-based method).
The effects of contraindications, upper age limit, referral patterns, quality of life and retransplantations are discussed below.

Contraindications

The potential impact of contraindications was estimated using medical data of referred, provisionally accepted and definitively accepted patients. Patients rejected because of age (mortality data are already age-specific) and those still under judgement were excluded from computations of the contraindication factor.
The disease-specific ratio of *definitively* accepted patients to *referred* patients was assumed to describe current selection policies[6]. The disease-specific ratio of *provisionally* accepted to *referred* patients was regarded to be an approximation of the effect of relaxed contraindications.
On average 75% of all referred patients within the age limits are rejected for contraindications, with relaxed contraindications this fraction will be about 60% (overall suitability 25% and 40% respectively, see Table 3). Notice that these are weighted averages: the rejection fractions for carcinoma's are 94% and 81%, for biliary atresia 56% and 16% respectively.

Upper age-limit

The Dutch protocol provides for an upper age-limit of 60 year. The influence of a higher upper age-limit is described by three sets of disease-specific values. They represent the increase of the number of deaths within a specific disease category, when 65 or 70 year would be the upper age limit, relative to the total number of deaths to 60 year (see Table 3)[23].

Table 2 **Deaths due to liver diseases amenable to liver transplantation. Upper age limit 60 year. Netherlands. 1979-1984.**

Disease category		ICD codes	Average annual number of deaths (abs)	Average annual number of deaths ($/10^6$)	Male/ female ratio	%<60 yr
I	Hepatocellular carcinoma	155.0	45.0	3.46	1.5	28%
II	Cirrhosis (excl PBC)	571.4	13.6	1.05	0.9	37%
III	Primary biliary cirrhosis	571.6	6.0	0.46	0.2	23%
IV	Unspecified chronic liver disease	571.5 571.9 571.8	64.5	0.50	2.0	24%
V	Biliary atresia	751.6	11.4	0.88	0.6	100%
VI	Primary sclerosing cholangitis	576.1	2.8	0.22	1.8	7%
VII	Metabolic diseases	270.2 277.1 271.0 277.4 275.0 277.6 275.1	11.3	0.87	1.5	71%
VIII	Other diseases	228.0 453.0 747.4	8.0	0.62	0.6	60%
Not included in mortality-based method						
IX	Acute hepatic failure	570.0	7.1	0.55	0.3	44%
X	Alcoholic liver disease	571.0 571.2 571.1 571.3	239.7	18.44	2.6	59%

Referral patterns

As discussed before, inequalities exist with regard to referral of patients from all provinces to the AZG. Accepting the referral pattern of the northern provinces as a standard, the referring level of the Netherlands is low. Half (145/274 ≈ 53%) of the expected patients is referred. Future improvement to 100% may be expected.

Table 3 Disease-specific adjustment factors to account for contraindications, quality-of-life considerations, age limit and retransplantation rate.

Disease category*	I HCC	II CIR	III PBC	IV UNSP	V BA	VI PSC	VII MTB	VIII OTHER
Referral								
- minimum level	0.53	0.53	0.53	0.53	0.53	0.53	0.53	0.53
- maximum level	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
No Contraindications / Suitability								
- minimum level**	0.06	0.41	0.60	0.41	0.44	0.40	0.26	0.26
- maximum level***	0.19	0.49	0.50	0.49	0.84	0.50	0.39	0.39
Quality-of-life								
- minimum level	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
- maximum level	1.00	1.10	1.10	1.10	1.00	1.00	1.10	1.10
Age-limit								
- 60 year	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
- 65 year	1.56	1.33	1.56	1.40	1.00	1.82	1.10	1.32
- 70 year	2.16	1.70	2.36	1.95	1.00	2.82	1.24	1.59
Retransplantation								
- minimum level	1.11	1.11	1.11	1.11	1.11	1.11	1.11	1.11
- maximum level	1.15	1.15	1.15	1.15	1.15	1.15	1.15	1.15

* categories cf. Table 3
** weighted average (all diagnoses): 0.25
*** weighted average (all diagnoses): 0.40

Quality-of-life

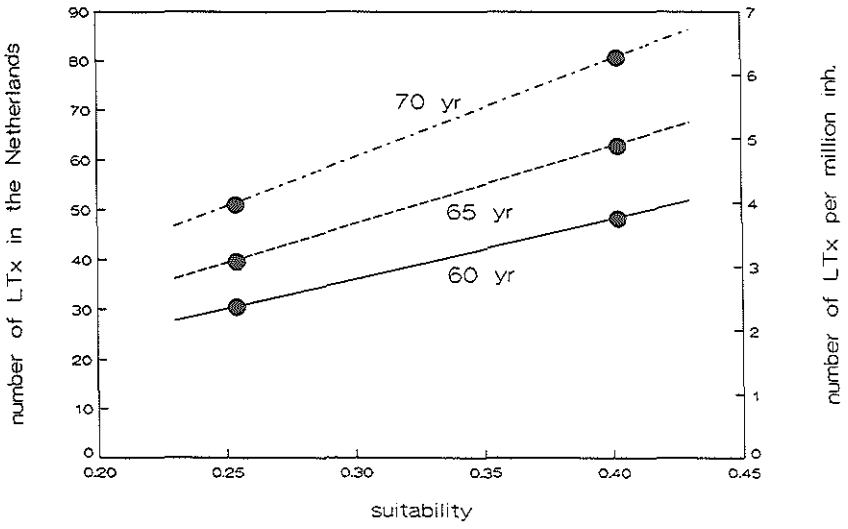
Analysis of the medical records of the transplanted patients revealed that occasionally improvement of the quality-of-life was the predominant reason for transplantation (intractable pruritus, intractable osteodystrophia)[7]. Though it may be argued that the presence of such complaints is strongly related to a poor prognosis, we assume that patients who were primarily transplanted for quality-of-life reasons are not represented by disease specific deaths in mortality statistics. Therefore we

assumed an independent disease-specific multiplication factor to account for patients accepted primarily for quality-of-life reasons. Two sets of values were chosen, one reflecting no and one reflecting modest impact of this aspect (see Table 3).

Retransplantation

Finally two retransplantation rates were chosen, 11% and 15% respectively (see practice-based method).

Figure 3 Mortality-based need for liver transplantation in the Netherlands. 15% retransplantations, quality-of-life transplantations included, 100% referral, no acute hepatic failure.



Final result

Combination of different levels of the adjustment factors allows for the presentation of many future scenario's of the total need for liver transplantation. In figure 3 the effects of age-limit and suitability (i.e. the effect of contraindications) are shown. The remaining adjustment factors all are assumed to have their maximum value. Choosing a lower level of these fixed adjustment factors will result in proportional lowering of the estimates. Note that acute hepatic failure was excluded.

3.3 Incidence-based method

Below a description of the results of is given of each diagnosis group separately following the disease classification of Table 2. Results are summarized in Table 4.

Hepato cellular carcinoma (HCC)

Only few reports exist with regard to the prevalence and incidence of this disease. Because the duration of illness is short and case-fatality rate is almost 100%, different measures of incidence and mortality should give comparable results. Vierling reported a yearly incidence of 10 to 70 per 10^6 males in an American standard population and half of this number for females[16].
Projection on the Dutch population results in 100 to 700 cases. The total number of admissions in 1982 in the Netherlands was 314. Mortality statistics revealed 162 deaths of whom 70 aged younger than 60. Suitability among HCC-patients may be derived from a study of 600 consecutive patients with HCC [24]. Only 98 could be regarded suitable for curative operation; 69 patients (12%) were in good clinical condition. This may serve as the upper level of the proportion of HCC-patients who might benefit from liver transplantation. Patients without apparent spread of disease often show recurrence of disease after curative treatment (either resection or transplantation) so a minimum level of about a third of 12% seems justified[25]. The AZG-experience shows a comparable low level of about 5% suitability (see Table 3).
Assuming 4 to 12% of the Vierling-projection suitable for liver transplantation results in an annual number of 4 to 85 patients. Starting with national mortality due to HCC of persons younger than 60 years results in a normative need of about 3 to 10 patients, which seems the most likely estimate. Thus an essentially mortality-based normative need resulted between 0.23 and 0.77 per 10^6 inhabitants.

Non-alcoholic cirrhosis (NAC)

This diagnosis category contains types of chronic cirrhosis of quite heterogeneous etiology. Most information is available about viral hepatitis which predominantly deals with the distribution of seroconversion due to viral hepatitis. The prevalence of hepatitis B carrier state, which is strongly related to the occurence of chronic cirrhosis, is supposed to be lesser than 1000 per 10^6 inhabitants of

the Western population[26]. However this information is of little help without a quantitative relation between serological evidence and clinical manifestations of the disease. Also little is known about prevalence and incidence of chronic cirrhosis of other etiology.Only Vierling provides for an estimate of the prevalence of all types of NAC (PBC excluded)[16]. Based on unknown sources a total number of 1800 - 9000 patients (all ages) per million inhabitants was projected for the United States. No information about the distribution of etiology and clinical stage of the disease was included.
Projection of this prevalence on the Dutch population results in about 23.000 to 117.000 patients. The SIG reported 675 (10% of the unspecified cirrhosis included) to 877 (25% of the unspecified cirrhosis included) hospital admissions of NAC in 1982 (all ages). We conclude that the American figures, if valid, do not seem realistic for the Dutch case. Dutch hospital data also do not yield a valid estimate because little is known about the re-admission rate and the severity of disease at admission.

Primary biliary cirrhosis (PBC)

We could dispose on two studies from the United Kingdom for an estimation of the frequency of PBC. Prevalence ranged from 40 to 54 per 10^6 inhabitants, incidence from 5.8 to 10.4 per 10^6 inhabitants[27,28]. The projected prevalence in the Netherlands is about 500 to 700 patients. In 1982 182 admissions were counted for this disease. As the prognosis of asymptomatic patients does not differ from healthy controls, knowledge about the distribution of disease states is mandatory for a valid estimation of the normative need for liver transplantation among PBC-patients[29]. Therefore the incidence-based method does not yield a valid estimate for primary biliary cirrhosis.

Biliary atresia (BA)

Vierling reported that biliary atresia occurred in about 0.6 to 1.0 of 10^5 living newborns [16]. Epidemiologic investigation of congenital diseases in the United Kingdom yielded an incidence of BA of about 0.8 per 10^5 living newborns[30]. Finally Houwen concluded form various epidemiological sources the incidence of BA in the Netherlands to be between the range of 0.4 to 0.7 per 10^5 living newborns[31].
The proportion of suitable candidates among patients with this congenital disorder is uncertain. Although primary liver transplantation may be a therapeutic option of the future, we assumed a maximum of 50% of newborns with BA to be a suitable candidate for liver transplantation.
About 175.000 living newborns are born in the Netherlands each year, 8 to 12 of whom have BA. Eight deaths with BA aged younger than 5 years were registered in 1982. This supports the abovementioned epidemiological data, provided that case-fatality was high before 1982, when children were not yet included in the liver transplantation-programme.
We conclude a normative need for liver transplantation after BA/Kasai of 4 to 6 each year, corresponding to 0.31 - 0.46 per 10^6 persons of a population with comparable reproduction rate.

Primary sclerosing cholangitis (PSC)

No reliable epidemiological data are available for this relatively new diagnostic entity[32]. The development of endoscopic techniques is responsible for the recent rapid increase of patients discovered.
Although PSC is associated with chronic inflammatory digestive disease (about 60%) no quantitative data are available of the proportion of with chronic inflammatory digestive disease with PSC[33].
Hospital-data showed 379 admissions with ICD-code 576.1. In 24 cases, 1 patient younger than 60

year, admission was followed by death due to the disease. National mortality statistics revealed on average 3 patients younger than 60 year dying in this disease category.
Because these results date from the time before wide-spread use of endoscopic devices we conclude that no incidence-based estimates for PSC are available, although AZG-experience suggests PSC is of minor importance (less than 0.10 per 10^6 inhabitants per year).

Metabolic diseases

A large number of different metabolic diseases affect the liver. Some of them have their primary locus within in the liver. liver transplantation is a therapeutic option in a considerable number of rare metabolic diseases but as a rule no population-based estimations are available on the frequency of their occurence.
Alfa-1-antitrypsin-deficiency (alfa-1-ATD) forms an exception, as population-based data are available from different sources. Genetically defined alfa-1-ATD occurs in 20 to 60 cases per 10^5 living newborns. Only 5 to 10% have clinical manifestations (cholestatic disease) and only 1 to 2% may benefit from liver transplantation (0.2 to 1.2 cases per 10^5 living newborns)[34].
Dick reported a comparable clinical incidence of alfa-1-ATD of 0.5 cases per 10^5 living newborns[30]. Projection on the annual Dutch birth cohort results in a normative need of 0 to 2 cases each year, corresponding to 0.03 - 0.08 per 10^6 persons of a population with comparable reproduction rate. Compared to alfa-1-ATD other metabolic diseases are of negligible quantitave importance.

Acute hepatic failure (AHF)

The clinical course of acute hepatic failure is short, usually resulting in either death or full recovery. Although AHF was excluded as an indication for liver transplantation during the period we investigated (1978-1987), recent developments show a considerable improvement of results, with subsequent incorporation of AHF into the AZG-protocol[18,19].
Two problems deserve special attention. First, no consensus exist with regard to the optimal therapeutic strategy, because of the prior unpredictability of those who will survive. Donor-availability is a second uncertainty in the decision-process. An agressive strategy certainly leads to treatment of patients, who otherwise would have been recovered. Expectant strategies lead to treatment of subacute hepatic failure only, with considerable loss of patients with an overwhelming course of their disease. Although it has been argued that auxiliary partial liver transplantation would be the surgical technique of choice, as a temporary support for those patients who recover, this outcome has not been observed yet. Most authors seem to agree on limited application of liver transplantation for AHF, for those who are "healthy enough to survive diagnostic check-up and donor-procurement and sick enough to be sure of a substantial probability of dying".
Epidemiological information on AHF is scarce. The number of 8.5 cases each year per 10^6 inhabitants of Vierling descend from an oral communication of Mosley and includes AHF of all causes (e.g. exacerbations of chronic infectious hepatitis) and all ages[35]. Comparison of this estimate with the actual admission rate in 1982 (117) gives similar results. Almost 20% died after admission, half of them being younger than 60 year. As the question remains which part of all admissions should be transplanted to prevent the major part of these deaths and which part of projected transplantations actually will be executed, the incidence-based method can not be used.

Table 4 Incidence-based estimation of the annual need for liver transplantation in the Netherlands. No retransplantation.

Disease category		Netherlands	per 10^6 inhabitants
I	Hepatocellular carcinoma	3 - 10	0.23 - 0.77
V	Biliary atresia	4 - 6	0.31 - 0.46
VI	Primary sclerosing cholangitis	0 - 1	0.00 - 0.08
VII	Metabolic diseases (alfa-1-ATD)	0 - 2	0.03 - 0.16

4 Discussion

We applied three methods to determine the need for liver transplantation. For those disease categories, where comparison was feasible we arrived at similar results (see Table 5). Given current selection criteria and 15% retransplantations, the annual normative need in the Netherlands ranges from about 25 to 70, corresponding to 2 to 5 per 10^6 persons.

Table 5 Annual number of expected transplants in the Netherlands for three diagnosis groups, 15% retransplantations included (number per 10^6 inhabitants between parentheses). Three methods.

Disease category	Practice-based	Mortality-based	Incidence-based
Non-alcoholic cirrhosis and PBC (II, III, IV)	15 - 37 (1.15 - 2.85)	14 - 17 (1.08 - 1.31)	NA
Acute hepatic failure (IX)	2 - 11 (0.15 - 0.84)	NA	NA
Other diseases (I, V, VI, VII, VIII)	8 - 19 (0.62 - 1.46)	16 - 32 (1.23 - 2.46)	7 - 19* (0.54 - 1.46)

* Disease category VIII assumed to be of neglegible importance.

Though results are roughly similar, this type of analysis may suffer from sources of uncorrectable bias and uncertainties. with regard to future development. Historical data may not account for future trends of the epidemiology of diseases (from the period 1978 - 1984). E.g. the incidence of hepatitis-B related chronic cirrhosis may be expected to fall due to the prevention of AIDS. For the larger part of the disease categories no short term change in incidence may be expected.
Three sources of uncertainty are related to changing inclusion and exclusion criteria. Increased effectivity of conventional therapy or the introduction of new therapeutic approaches may decrease the need for liver transplantation. Examples are various drug regimes for hepatitis-B, prostaglandines for acute hepatic failure, ursodeoxycholic acid for the treatment of PBC and PSC[36,37]. Application of liver cell transplantation may not be expected in the next future and will probably only affect the need for some minor diagnostic categories (metabolic diseases). At the other side improved results of liver transplantation in acute hepatic failure and formerly too ill regarded patients will increase the need for liver transplantation. For short term forecasts these are neglible compared to the impact of changing the age limits or indicationa (alcohic cirrhosis), see Table 2 and 3.

In addition to these general remarks some method-specific questions remain. Though the practice-based method accounts for regional differences of referral practices, it assumes complete referral in the northern provinces. This assumption seems valid as the AZG is the only hepatological centre in the northern and adjacent provinces, with traditionally complete adherence of patients.
Questions may remain with regard to the validity of the use of mortality data with an upper age limit of 60 year (mainly for non-alcoholic and primary biliary cirrhosis). Their use assumes a neglegible time interval between supposed application of the medical procedure and death due to the specific disease. The value of the 65-year adjustment factor for non-alcoholic and primary biliary cirrhosis may be used as an indicator for the magnitude of bias when transplantation would take place on average 5 year before death. After intrapolation a time lag of 1 year turned out to result in an underestimation of need for non-alcoholic and primary biliary cirrhosis of 5% to 10%. Coding errors are another source of uncertainty. If 25% instead of 10% of unspecified cirrhosis is supposed to be non-alcoholic, normative need raises about 15%; 0% leads to a decrease of 11%.

Provided that gross alteration of inclusion criteria does not occur, our estimations seem valid enough to support governmental decisions for the next few

years. The number of liver transplantation, which actually will be performed depends on the availability of resources. Lack of donor organs generally (children excluded) may not be expected in the next few years[38]. In absence of financial constraints only one liver transplantation centre with a capacity of 50 transplantations annually will suffice in the Netherlands in the near future. Our study showed the existence of considerable regional differences in referral pattern. Patient's negative attitude towards high technology medicine, relative unawareness of the current effectivity of liver transplantation and unwillingness of specialists to refer patients to another hospital contributed to this undesirable state of affairs. Governmental strategies for optimizing application of liver transplantation should pay attention to this problem.

Lastly, the application of our results in other countries is of interest. Though various transplant centres have published expectations of the need for liver transplantation, only O'Grady published an empirical estimation for chronic liver diseases and acute hepatic failure, using British mortality data[25]. Obviously some minor disease-categories were excluded. About 30% of those who died from chronic liver disease were thought to be suitable for liver transplantation (1000 out of 2937), which is close to our estimates. However, O'Grady's ultimate normative need ranges from 4 to 20 per 10^6 persons, which far exceeds our estimates. Apparently different inclusion criteria are responsible for this discrepancy. O'Grady included all deaths due to cirrhosis regardless of age and diagnosis. Only about 15% of them meet both our age criterion (age below 60) and diagnosis criterion (no alcoholic cirrhosis), which explains the magnitude of the difference.

Though evidence is limited, we believe our results are internationally applicable for disease categories without known causal origin. Alcoholic cirrhosis, hepatocellular carcinoma and viral hepatitis with its chronic sequelae show large international differences [13, 14, 15]. For most countries our estimates will be a serious underestimation when these disease-categories are included.

References

1. Cuthbert JA. Southwestern Internal Medicine Conference: hepatic transplantation. Am J Med Sciences 1986; 291:286-296.
2. Chu F, Cotter D, Hamilton MAT. First-year report on adult liver transplantation operating costs, medicare payments and utilization rates. Medical Technology and Practice Patterns Institute, Inc. Washington DC 1988.
3. Kankaanpää J. The cost-effectiveness of liver transplantation. University of Texas (Houston), Thesis MPH 1988.
4. National Institutes of Health Consensus Development Conference Statement: Liver transplantation. June 20-23, 1983. Hepatology 1984; 4:107S-110S.
5. Bonsel GJ, Habbema JDF, Bot ML, et al. A technology assessment of liver transplantation. An evaluation of the Groningen liver transplant programme 1977 - 1987 (in Dutch). Ned T Geneeskd 1989, accepted for publication.
6. Gips CH, Krom RAF. Indications and procedure for referral of potential candidates for liver transplantation (in Dutch). Ned T Geneeskd 1981; 125:875-878.
7. Maddrey WC, Thiel DH van. Liver transplantation: an overview. Hepatology 1988; 8:948-959.
8. Bradshaw J. A taxonomy of social need. In: McLachlan G. (ed.). Problems and progress in medical care, 7th series, London, Oxford University Press 1972.
9. Williams A. "Need" - an economic exegesis. In: Culyer AJ, Wright KG (ed.). Economic aspects of health services. York, Martin Robertson 1978.
10. Evans RW, Manninen DL, Overcast TD, et al. The National Heart Transplantation Study: Final Report. Seattle, Washington: Battelle Human Affairs Research Centers, 1984.
11. Buxton M, Acheson R, Caine N, et al. Costs and benefits of the heart transplant programmes at Harefield and Papworth Hospitals: Final Report. Department of Economics, Brunel University and Department of Community Medicine, University of Cambridge, 1985.
12. O'Grady JG, Williams R. Present position of liver transplantation and its impact on hepatological practice. Gut 1988; 29:566-570.
13. Steiner PE. World problem in the cirrhotic diseases of the liver. Trop Geogr Med 1964; 16:175-205.
14. Massé L, Juillan JM, Chisloup A. Trends in mortality from cirrhosis of the liver, 1950-1971. World Health Statistics Report 1976; 29:40-67.
15. Holland WW (ed.). European Community atlas of "avoidable death". Oxford University Press 1988.
16. Vierling JM. Epidemiology and clinical course of liver disease: identification of candidates for hepatic transplantation. Hepatology 1984; 4:84S-94S.
17. Williams R, Gimson AES. An assessment of orthotopic liver transplantation in acute liver failure. Hepatology 1984; 4:22S-28S.
18. Editorial. Transplantation for acute liver failure. Lancet 1987;ii: 1248-1249.
19. Bismuth H, Samuel D, Gugenheim J, et al. Emergency liver transplantation for fulminant hepatitis. Ann Int Med 1987; 107:337-341.
20. Iwatsuki S, Esquivel CO, Gordon RD, et al. Liver transplantation for fulminant hepatic failure. Semin Liver Dis 1985; 5:325-328.

21. Bismuth H, Ericzon BG, Rolles K, et al. Hepatic transplantation in Europe: first report of the European Liver Transplant Registry. Lancet 1987; ii:674-676.
22. World Health Organization. Manual of the International Statistical Classification of Diseases, Injuries and Causes of Death. 1975 Revision. Volume 1. Geneva, 1977.
23. Starzl TE, Todo S, Gordon R, et al. Liver transplantation in older patients. N Engl J Med 1987; 316:484-485.
24. Okuda K, Obata H, Nakajima Y, et al. Prognosis of primary hepatocellular carcinoma. Hepatology 1984; 4:3S-6S.
25. O'Grady JG, Polson RJ, Rolles K, et al. Liver transplantation for malignant disease. Results in 93 consecutive persons. Ann Surg 1988; 207:373-379.
26. Miller DL, Farmer RDT. Epidemiology of diseases. Oxford, Blackwell 1982.
27. Triger DR. Primary biliary cirrhosis: an epidemiological study. Br Med J 1980; 281:772-775.
28. Hislop WS. Primary biliary cirrhosis: an epidemiological study (letter to the editor). Br Med J 1980; 281:1069-1070.
29. Roll J, Boyer JL, Barry D, et al. The prognostic importance of clinical and histologic features in asymptomatic and symptomatic primary biliary cirrhosis. N Engl J Med 1983; 308:1-7.
30. Dick MC, Mowat AP. Hepatitis syndrome in infancy - an epidemiological survey with 10 year follow up. Arch Disease Childhood 1985; 60:512-516.
31. Houwen RHJ, Zwierstra RP, Severijnen RSVM, et al. Extrahepatic biliary atresia in the Netherlands. The longterm prognosis after surgery. Arch Diseases Childhood 1989; 64:214-218.
32. Dickson ER, LaRusso NF, Wiesner RH. Primary sclerosing cholangitis. Hepatology 1984; 4:33S-35S.
33. Larusso NF, Wiesner RH, Ludwig J, et al. Primary sclerosing cholangitis. N Engl J Med 1984; 310:899-903.
34. Dutch Health Council (Medical Advisory Board of the Dutch Government). Recommendations on auxiliary partial liver transplantation (in Dutch). The Hague 1985.
35. Berk PD, Popper H. Fulminant hepatic failure. Am J Gastroenterology 1978; 69:349-400.
36. Aach RD. Treatment of chronic type B viral hepatitis. Ann Intern Med 1988; 109:89-91.
37. Poupon R, Poupon RE, Calmuus I, et al. Is ursodeoxychoilic acid an effective treatment for primary biliary cirrhosis ? Lancet 1987; i:834-836.
38. Charro FT de, Bonsel GJ, Bot ML. Cost and effects of liver transplantation, vol. 6: The availability of donor livers (in Dutch). Department of Economics, Faculty of Law, Erasmus University Rotterdam, 1988.

Medical Technology Assessment. The price of advice.

Gouke J. Bonsel
Frans F.H. Rutten

Appeared in Dutch in Medisch Contact, 16 (1990); 512-516
Published with permission of the editor

Whilst politicians are still arguing over the selection criteria for heart transplantation and while they are postponing a policy decision on liver transplantation until at least 1992, at the same time being trapped in the license policy concerning in vitro fertilization, the investigators look back in this article at the three evaluation studies, that the same politicians wanted to have performed for these decisions[1,2,3]. The reason for this retrospective is not a finished decision process, although the availability of the outcome of the study in 1988 would have implied a firm decision. The direct cause is the appearance of a report in January 1990, written by Van Rossum under the authority of the National Health Insurance Board[4]. In this report the value of these three studies - usually indicated by Medical Technology Assessments - has been investigated from the client's point of view. This article describes the view of the institute for Medical Technology Assessment (iMTA) on this topic. The paper is structured by six questions pertinent to the issue: What are the criteria one should apply at an evaluation of evaluations? What has been the relevance of the evaluations at the political level? Have the studies produced new insights? What is the spin-off of the studies? Is a cheaper option available? Are there alternatives anyway?

1. Criteria

The selection of the criteria to judge the value of the evaluation studies is, obviously, crucial. The value of the studies as perceived by their users has been selected as the primary criterion in the report-Van Rossum. This value was determined by interviewing the relevant decision-makers in the summer of 1989, asking them to adress the following question: what was, according to your experience, the role of the evaluation studies in the advising and provisional decision-making on the inclusion of the evaluated interventions in the insurance package? It may be argued that the actual or the pretended utilization of information by health politicians depends on more than on the quality of information. Therefore we think a second criterion has to be applied if we evaluate these evaluations, viz. the question if the information, which was asked for by the clients, has really become available. That question to the *value of information* hardly comes up in the previously mentioned report. However, an essential condition for rational policy choices is the availability of objective and reliable information.

Whether such information will actually be used depends ultimately on the rationality of this process of choice itself.
Thus, a specific definition of what valuable information from a medical technology assessment consists of, is necessary. In our opinion, this definition includes a.o. that information should be non-distorted, reliable, unweighted, comprehensive and ready available at the time it is needed. This listing looks trivial, but we will illustrate that the search for suitable information as defined above (criterion 2) and the promotion of the utilization of these studies in health policy decisions (criterion 1) may be at conflict.
The report-Van Rossum makes some recommendations for promoting utilization (criterion 1): the involvement of the health policy makers in the initiation, design and execution of an MTA-investigation should be even greater than it already was in the three MTA-studies mentioned above; a small distance between policy maker and investigators should be the aim. In the extreme case that would imply that the research team functions within the policy body. These recommendations of Van Rossum follow from the dominant position of the user in his definition of the value of information: this value was operationalized as the actual utilization as perceived by those who instituted the MTA. To fulfil the demands of good information (criterion 2) some other conditions has to be satisfied. Three of them are:
1. the assignment of an independent role to the investigator in designing the study;
2. the opportunity to collect primary data, if secondary sources appears to be insufficient;
3. the execution of the study within an universitary setting, which implies political independence of the research team.

Ad 1. The separate responsibility of MTA-investigators with regard to the definition of study objectives and study design should be acknowledged. This responsibility turns into a major role on some occasions. Two examples are given.

In the transplantation studies we remember lengthy discussions about the necessity of estimating survival without transplantation and the necessity of estimating the quality of life changes.

In a cost-effectiveness analysis of a new cholesterol lowering drug under authorization of a pharmaceutical industry, some time had to be spent discussing the fact whether or not to include certain (standard) costs and effects in the analysis.

Ad 2. Another two examples illustrate problems with respect to the second condition.

The relation between the research team at the IVF study and the secretariat of the NHIB initially provoked problems with regard to the data collection: at the start of the IVF study few observations were known on male subfertility. This was a registration-artifact due to the fact that this indication was not qualified by the NHIB for subsidy of the participating IVF centres.

Data collection for cost registration in all studies was influenced demonstrably by the presumed application of cost results at tariff negotiations.

Ad 3. The scientific independency of the MTA team could be endangered if policy goals guide the investigations too much. Independency is necessary, be the client a governmental agency, or a pharmaceutical company. Consequently, free publication by the MTA-team is a key instrument to establish their independence. If this freedom leads to impassable objections from the side of the client or the investigated patients, the study may be withdrawn.

In the case of the Dutch MTA's the investigators think all the abovementioned conditions were satisfied, hence there is no reason to recommend, like Van Rossum, a shift in responsibilities between MTA-investigators and policy body to attain another balance between conditions for good research and conditions for the utilization of results.

We agree with the report-Van Rossum that in some cases the actual use of information by health policy makers has been to some extent insufficient. We might think of their rather slow reaction and of the occasional presence of a non-decision policy. Three examples may illustrate the judgment.

The large inter-hospital variation of in vitro fertilization results was an excellent pretext for the license policy of Article 18 of the Hospital Provisions Act; despite this variation, health policy failed to device a rational regulation, as been memorized recently by Rigter[5].

In heart transplantation age limits are currently employed, which certainly do not back up the results of our MTA.

Finally, liver transplantation is still not regularly compensated to children with biliary atresia, although the results of liver transplantation indicate transplantation to be the optimal option available. Paradoxically, no continued research was considered essential for the decision making on transplantation in these children (contrarily to the adults).

2. Policy relevance

Have the evaluations been used for policy-making? This question is thoroughly answered in the report-Van Rossum. The results appear to be applied rather quickly and extensively by both the National Health Insurance Board and the Health Council.

In our experience, the health politicians sometimes disliked the obvious lack of pre-arranged decisions in our reports. Van Rossum also considers this to be a disadvantage; he speaks of 'solely indirect use' from the advisory boards and the health care authorities. However, in our opinion this 'disadvantage' is the result of an intentional separation of responsibilities. Following this intention, the apparent absence of other information (for example on ethical aspects) is no real inadequacy, because the client explicitly provided his need for that information in another way. This deliberate efficiency of information generation succeeded as can be illustrated in the case of heart transplantation.

In the medical technology assessment of heart transplantation the socalled scenario analysis clearly showed that for optimal utilization of all donor hearts offered a large waiting list would be profitable. However, the price for this optimal utilization is a considerable number of patients dying on the waiting list - not quite equivalent with the situation that the same patients dy unaware of being a candidate for transplantation - and a grinding selection issue. The ethical aspects of such problems indeed were thoroughly discussed by the Health Council. No duplication of research efforts happened[2-6].

3. New information

The question is relevant whether these studies have contributed new information as, in the clinicians perception, they have been conducted by relative outsiders. Although the answer is not the single criterion to measure the quality or the benefit of the research the answer is important as a justification towards the clinical investigators. In our opinion, information is new, if it 1. considerably differs from present scientific knowledge, 2. it fills up an information void, 3. it restructures present and/or new knowledge. Two dates are important in this respect: the starting (summer 1985) and the finishing date of the studies (summer 1988).

In general, one may state safely that at the completion of the studies in 1988 details of the costs, the effectiveness and the throughput of the programme were unavailable elsewhere. The cost-effectiveness analysis, which we presented, and, in the case of heart transplantation, the detailed future forecasting (scenario's), offered a new structure to the information available.

If the focus is on clinically knowledge, especially on the effectiveness of transplantation, our experience is interesting.

In the case of *heart transplantation* the MTA results didn't differ from a priori expectations and from the publications of one of the centres involved. This was maybe due to the fact that in all centres one protocol was used and, apparently, care of the same high quality was supplied.

The picture was completely different in the *liver transplantation* and in the vitro fertilization study. The one- and five year survival probabilities after liver transplantation were, according to literature, approximately 80% and 75% in 1985 (in 1988 even higher), but in the Dutch programme these probabilities appeared to be lower, 68% and 58% respectively. Not a bad clinical performance, but inferior statistics (uncontrollable 'case-selection, combination of children and adults, and other forms of publication bias) explained these differences. The MTA-estimates of expected survival without transplantation (all forms of cirrhosis: one year survival 51%) was to such an extent different from clinical expectations (cirrhosis: at last three months survival, one year survival < 10%) that many people doubted the validity of the MTA-results. Only when the two biggest transplantation centres in the world published a similar analysis in the New England Journal of Medicine and now that the Dutch analyses has appeared in the Lancet, more people accept the MTA-results on effectiveness of liver transplantation[7,8]. Results on other aspects of liver transplantation also showed large discrepancies between scientific sources of information and our empirical MTA-report. Two examples of discrepancies are given, starting with the costs of liver transplantation. The business economics department of the Academic Hospital Groningen estimated these costs more than twice higher than we later calculated. The costs of liver transplantation in the U.S. as they have appeared in several reports, are also much higher, even if we assume that one Dutch florin spent at health care in the Netherlands corresponds with one US dollar spent at health care in the U.S. Secondly, the need for liver transplantation in the American literature is estimated at 20-250/10^6 inhibitors, in the English literature at 4-20/10^6 and in the Netherlands at 2-5/10^6 inhibitors.

Lastly, we shall present some evidence of the *in vitro fertilization* study. At the start of the MTA, success rates of 20% to 30% per in vitro fertilization treatment were reported. In vitro fertilization appeared to be clearly superior over tuba surgery. The advisory committee on IVF of the NHIB, proposed to accept only those hospitals for participation in the MTA only, which could show at least a 10% success rate per IVF-procedure. This criterion apparently was generous, considering the prior information about success rates. Fortunately that criterium was not effectuated: in the long run only two out of five participating hospitals did show a success rate over 10%, whilst two out of five hospitals were just below and one was far below the percentage. At this moment, one of the centres communicates through the newspaper that the effectiveness of IVF treatments could at least double by application of a new method of hormone stimulation. In our view only primary data collection in a controlled way can prove this bold prediction.

According to the report-Van Rossum the studies were not valid, partly because of the fact that no randomized clinical trial was performed. We disagree with this global judgment, though some research issues were particularly difficult to address due to pathological complexities (for example the prediction of survival of patients suffering from heart failure) or due to the fact that a research effort was acquired that was far beyond this medical technology assessment (for example the determination of the number of persons in the Netherlands with severe heart failure of severe liver dysfunction).

It has been proved though, that the studies can stand the comparison with cost-effectiveness studies abroad. This is found to be true according to the resulting publications, but also according to the requests transfer of the applied methodology to other places in the Netherlands and abroad.

Nevertheless, we are aware of the fact that further development and, especially, standardization of the applied methodology is necessary[9]. Lack of standardization is partly responsible for the above mentioned differences in results. Consensus on a uniform protocol for the execution of medical technology assessments is urgently needed. Thus the results of various studies will become comparable and they will consequently constitute a better point of action for policy-making.

4. Spin-off

The studies clearly produced considerable spin-off.

First, the impact of the studies for settling the expertise in this area in the Netherlands cannot be underestimated. This effect may well be to the credit of the Health Insurance Executive Board, the Health Council and the Ministry of Welfare, Public Health and Cultural Affairs. These agencies were at the start of the studies and committed themselves to take account of the results in their policy-making. After the assignments of the MTA to the centres in Rotterdam and Maastricht, these centres acquired a lot of new contracting research, sometimes after competition with reputable institutes in the U.K. and the U.S. This development has given rise to a national centre for MTA, which is supported by the parties which conducted the evaluation studies. Moreover, the MTA-investigators established an informal communication network, to which many international MTA-centres participate.

A second spin-off effect is the increased knowledge on methods, possibilities and limitations of medical technology assessment, and of problems concerning its interpretation. This side effect is visible in both health politicians and the investigators. The effect results from their intensive communication. The Dutch policy-maker is at the frontier of MTA, at least in Europe. This position contributes to the responsible application of MTA research.

Thirdly, we denominate the increased interest for the methods of cost-effectiveness analysis, at least by some clinicians. This interest is of significance as the simultaneous development of a clinical research programme and its associated MTA offers the best opportunity for a cost-effectiveness analysis. Regarding the developmental medicine* it is important that clinical investigators are aware of the relatively low marginal costs of performing a complete evaluation study instead of an effectiveness study only. The future must prove whether the dangers of publication bias and the heterogeneity of methods for the determination of effects and costs will negatively influence the results of developmental medicine studies.

---

* In the Netherlands the National Health Insurance Board has insituted a fund ('Ontwikkelingsgeneeskunde', tr.: Developmental medicine) to support the new medical technologies at the stage they are introduced for clinical use. The application of a grant depends a.o. on the inclusion of a formal evaluation study in the design.

5. Is there a cheaper option?

From an absolute point of view, the MTA-studies have been rather expensive: the costs ranged from Hfl. 1,5 million (IVF) to Hfl. 1,9 million (heart transplantation[4]) (total costs include the data management of the participating institutes!). It can be cheaper on next occasions, of course. The experience with the design and the execution of empirical research in this area will lead to a more efficient approach in future. For example, the already developed instruments regarding quality of life and utility measurement can be applied in new studies.

Whether an alternative exists, which is indeed cheaper, depends on the question whether primary data collection is required and to which degree information is already available in hospitals or outside.

Sometimes primary data collection is *not possible* or even *unnecessary*.

Data collection is difficult if a new drug still has to be or just has been registered. Usually, available time for investigation meaningful primary data collection according to scientific standards. A modelling approach, like we adopted for the evaluation of the new drug simvastatine is more appropriate in that case. More time for evaluation would result if the onus of proof rests with the producer. After all, the producer is responsible for the developments at an early stage, when data on its security and its effectiveness are collected.

Secondly, primary data collection is unnecessary if a meta analysis of published studies provides enough evidence on the effectiveness of a technology and if the estimation of the costs in the Netherlands is possible by adaption of costs figures from abroad. However, one must be aware of publication bias and unduly optimism in first papers on innovations. Our institute is involved in the meta-analysis of several issues. For example the transfer of the Dutch model for heart transplantation evaluation to the Spanish situation and, at an earlier stage, the application of the methods and results the American and the English studies on heart transplantation in the Netherlands. Sofar the evidence is somewhat equivocal.

Ample knowledge of and long experience with MTA research appears to be necessary to conclude a 'quick and dirty' medical technology assessment successfully, particularly if, unlike the heart transplanation case, predecessors on the topic are absent. Particularly the estimation of the global magnitude of effects in absence of empirical data is demanding.

If primary data collection and subsequent analysis *is* indeed *necessary,* one its costs should be related to the expenses of the evaluated technology. If we consider the expenses of in vitro fertilization over five years, only 1% of the expenses has been applied for evaluation. For the liver transplantation study this is approximately 4%.

6. Are there alternatives any how?

Evaluation of new medical technologies may take various forms. In the Netherlands the calling in of the Health Council often appears to be the best option available. The increasing professionalization of this unique institute and the continuing loyal cooperation of medical experts guarantees a useful answer on many occasions.

Another form is the institution of a project on developmental medicine, which should be the responsibility of the clinical experts themselves. Future will show which balance between the information from such research and that of an independent technology assessment should be preferred.

Some people, amongst who Van Rossum, have claimed that a technology assessment should enhance more than data collection only. They suggest that external steering of medical technologies is feasible by means of the utilization of existing information and by the application of societal criteria (constructive TA). Thus, a constructive technology assessment not only requires that the technical aspects are known at an early stage of the development of a technology, but it also considers the social aspects of the introduction of the technology. Three remarks given.

Developmental medicine could, in some cases, be considered as a constructive medical technology assessment, particularly if the obtained insights improve the performance of a new technology or the circumstances under which these are performed. We think the relevance of constructive technology assessment depends on the type of technology and its developmental stage.

In general the relevance should not be overestimated. Many medical technologies are introduced from outside our country, for example new drugs and new medical equipments. Constructive MTA is not useful to evaluate these mature technologies. Policy-makers, who are confronted with these technologies, have only to make a rather simple decision about allowance on the market and subsequent coverage by health insurance companies. Therefore, a constructive TA is no

alternative for the type of MTA-studies which are finished now. Our previous remark applies i.e. the shift of the onus of proving the effectiveness of a product from the health care authorities to the producer. This policy change probably will influence the development of a product by a producer in a way resembling the presumed influence of constructive technology assessment.

Anyway, constructive technology assessment assumes that the technological development by the medical profession may be guided externally, an assumption which has been challenged[10].

7. Conclusion

We conclude with the following remarks.
First, obviously the report-Van Rossum did not address some important questions. Secondly, after re-reading the report, one is left with too much a feeling of the infallible excellence of our advisory bodies. More distance from the evaluator to the client would possibly have made the report as tempting for him as it was for us. Finally, we sincerely hope that the atmosphere of collaboration, which has been built up with clinicians in the past, will be continued in the future, be it in the context of studies of developmental medicine or, in the context of MTA studies initiated by policy-makers on any level in health care.

8. References

1. Charro FTh de, Bonsel GJ, Hout BA van. The costs and effects of heart transplantation. Final report (in Dutch). Rotterdam: institute for Public Health and Social Medicine, Erasmus University Rotterdam, 1988.

2. Bonsel GJ, Habbema JDF, Bot ML, et al. A technology assessment of liver transplantation: a research to the liver transplantation programme of Groningen of 12977-1987 (in Dutch). Ned Tijdschr Geneeskd 1989: 133; 1406-14.

3. Haan G. Effects and costs of in vitro fertilization (in Dutch). Thesis at the University of Limburg, Maastricht, 1989.

4. Rossum W van. Decision-making and medical technological aspects research: an analysis of the utilization value of TA-studies to heart and liver transplantations and in vitro fertilization (in Dutch). Report. Goningen: Faculty of business economics, department of organization, Technology and Innovation/working group Scientific Research, University of Groningen, 1989.

5. Rigter H. The social sight of medicine: concerning the development of medical technologies in this century (in Dutch). Declamation. Rotterdam: Faculty of Social Science and Medicine, Erasmus University Rotterdam, 1989.

6. Rigter H. Who's next? About selection and waiting lists in health care (in Dutch). Backing study. Den Hague: Health Council, 1989.

7. Markus BH, Dickson ER, Grambsch P, et al. Efficacy of liver transplantation in patients with primary biliary cirrhosis. N Eng J Med 1989: 320; 1709-134.

8. Bonsel GJ, Klompmaker IJ, Veer F van 't, et al. Liver transplantation prolongs long term survival in patients with primary biliary cirrhosis. Lancet (1990): 335; 493-97.

9. Habbema JDF, Casparie AF, Mulder JH, Rutten FFH. Medical technology assessment and health care policy (in Dutch). Alphen aan den Rijn: Samson Staffleu, 1989.

10. Mintzberg H. The structuring of organizations. Englewood Cliffs: Prentice Hall, 1979.

LIST OF CO-AUTHORS

Frank T. de Charro, Department of Economics, Faculty of Law, Erasmus University Rotterdam, The Netherlands

Marie-Louïse Essink-Bot, Department of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands

The EuroQol Group (see appendix 8)

J. Dik F. Habbema, Department of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands

Ben A. van Hout, institute for Medical Technology Assessment, Department of Health Care Policy and Management, formerly: Department of Economics, Faculty of Law, Erasmus University Rotterdam, The Netherlands

Ids J. Klompmaker, Liver Transplant Group, Academic Hospital Groningen, The Netherlands

Paul J. van der Maas, Department of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands

Frans F.H. Rutten, institute for Medical Technology Assessment, Department of Health Care Policy and Management, Erasmus University Rotterdam, The Netherlands

Wim E. Saris, Sociometric Research Foundation, Amsterdam, The Netherlands

Maarten J.H. Slooff, Liver Transplant Group, Academic Hospital Groningen, The Netherlands

Fred van 't Veer, formerly: Department of Public Health and Social Medicine, Erasmus University Rotterdam, The Netherlands

Martin Verweij, formerly: Sociometric Research Foundation, Amsterdam, The Netherlands

CURRICULUM VITAE

Opleiding

1956	Geboren te Beverwijk
1968 - 1974	Gymnasium B te Alkmaar
1974 - 1982	Geneeskunde, Medische Faculteit, Vrije Universiteit Amsterdam (cum laude)
1975 - 1977	Filosofie, Centrale Interfaculteit, Vrije Universiteit Amsterdam
1985 - 1987	Sociale wetenschappen, Faculteit der Sociale Wetenschappen, Rijksuniversiteit Utrecht
1986 - 1990	Sociale Geneeskunde, Medische Faculteit, Erasmus Universiteit Rotterdam (opleider: Prof.dr P.J. van der Maas)

Werk

1982 - 1983	Docent ziekenhuis (Interne Geneeskunde) Docent sociale academie (Medische Informatie) Docent school (piano)
1983 - 1985	Medisch adviseur, Ziekenfondsraad
1985 - 1988	Arts-onderzoeker, Instituut Maatschappelijke Gezondheidszorg, Erasmus Universiteit Rotterdam
1988 -	Detachering instituut voor Medische Technology Assessment, Erasmus Universiteit Rotterdam

The investigations, which constitute the basis of this thesis, have been executed at the Institute for Public Health and Social Medicine, Erasmus University Rotterdam, the Netherlands.
The studies of liver transplantation and heart transplantation were supported by grants of the National Health Insurance Board.
The participation in the EuroQol-group was partially supported by a grant of the Dutch Fund for the Promotion of Health Services Research.

De onderzoeken, die ten grondslag liggen aan dit proefschrift, werden uitgevoerd bij het Instituut Maatschappelijke Gezondheidszorg van de Erasmus Universiteit te Rotterdam.
De hier beschreven studies naar lever- en harttransplantatie werden gesubsidieerd door de Ziekenfondsraad.
De bijdrage aan de EuroQol-studiegroep werd mede mogelijk gemaakt door een subsidie van het Stimuleringsfonds voor Gezondheidszorg Onderzoek.